This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

#7 Attachment
10/023586

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 July 2001 (12.07.2001)

PCT

(10) International Publication Number
WO 01/49847 A2

(51) International Patent Classification[7]: C12N 15/12, 15/62, 15/11, C07K 14/72, 19/00, 16/28, C12Q 1/68, G01N 33/50, 33/566, A61K 38/17

(21) International Application Number: PCT/US00/35309

(22) International Filing Date:
22 December 2000 (22.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/475,790 30 December 1999 (30.12.1999) US

(71) Applicant *(for all designated States except US)*: **MILLENNIUM PHARMACEUTICALS, INC.** [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **GLUCKSMANN, Maria, Alexandra** [AR/US]; 33 Summit Road, Lexington, MA 02173 (US). **WHITE, David** [US/US]; 35 Hollingsworth Avenue, Braintree, MA 02184 (US).

(74) Agents: **COULTER, Kathryn, L.** et al.; Alston & Bird LLP, P.O. Drawer 34009, Charlotte, NC 28234-4009 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/49847 A2

(54) Title: 26904, 38911, AND 39404, NOVEL SEVEN-TRANSMEMBRANE PROTEINS/G-PROTEIN COUPLED RECEPTORS

(57) Abstract: The present invention relates to newly identified seven-transmembrane proteins, including proteins that function as receptors belonging to the superfamily of G-protein-coupled receptors. The invention also relates to polynucleotides encoding the seven-transmembrane proteins/receptors. The invention further relates to methods using the seven-transmembrane protein/receptor polypeptides and polynucleotides as a target for diagnosis and treatment in seven-transmembrane protein/receptor-mediated and related disorders. The invention further relates to drug-screening methods using the seven-transmembrane protein/receptor polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the seven-transmembrane protein/receptor polypeptides and polynucleotides. The invention further relates to procedures for producing the receptor polypeptides and polynucleotides.

26904, 38911, AND 39404, NOVEL SEVEN-TRANSMEMBRANE PROTEINS/G-PROTEIN COUPLED RECEPTORS

FIELD OF THE INVENTION

The present invention relates to newly identified seven-transmembrane proteins, including proteins that function as receptors belonging to the superfamily of G-protein-coupled receptors. The invention also relates to polynucleotides encoding the seven-transmembrane proteins/receptors. The invention further relates to methods using the seven-transmembrane protein/receptor polypeptides and polynucleotides as a target for diagnosis and treatment in seven-transmembrane protein/receptor-mediated and related disorders. The invention further relates to drug-screening methods using the seven-transmembrane protein/receptor polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the seven-transmembrane protein/receptor polypeptides and polynucleotides. The invention further relates to procedures for producing the seven-transmembrane/receptor polypeptides and polynucleotides.

BACKGROUND OF THE INVENTION

G-protein coupled receptors

G-protein coupled receptors (GPCRs) constitute a major class of proteins responsible for transducing a signal within a cell. GPCRs have three structural domains: an amino terminal extracellular domain, a transmembrane domain containing seven transmembrane segments, three extracellular loops, and three intracellular loops, and a carboxy terminal intracellular domain. Upon binding of a ligand to an extracellular portion of a GPCR, a signal is transduced within the cell that results in a change in a biological or physiological property of the cell. GPCRs, along with G-proteins and effectors (intracellular enzymes and channels

modulated by G-proteins), are the components of a modular signaling system that connects the state of intracellular second messengers to extracellular inputs.

GPCR genes and gene-products are potential causative agents of disease (Spiegel *et al., J. Clin. Invest. 92*:1119-1125 (1993); McKusick *et al., J. Med. Genet. 30*:1-26 (1993)). Specific defects in the rhodopsin gene and the V2 vasopressin receptor gene have been shown to cause various forms of retinitis pigmentosum (Nathans *et al., Annu. Rev. Genet. 26*:403-424 (1992)), and nephrogenic diabetes insipidus (Holtzman *et al., Hum. Mol. Genet. 2*:1201-1204 (1993)). These receptors are of critical importance to both the central nervous system and peripheral physiological processes. Evolutionary analyses suggest that the ancestor of these proteins originally developed in concert with complex body plans and nervous systems.

The GPCR protein superfamily can be divided into five families: Family I, receptors typified by rhodopsin and the β2-adrenergic receptor and currently represented by over 200 unique members (Dohlman *et al., Annu. Rev. Biochem. 60*:653-688 (1991)); Family II, the parathyroid hormone/calcitonin/secretin receptor family (Juppner *et al., Science 254*:1024-1026 (1991); Lin *et al., Science 254*:1022-1024 (1991)); Family III, the metabotropic glutamate receptor family (Nakanishi, *Science 258 597*:603 (1992)); Family IV, the cAMP receptor family, important in the chemotaxis and development of *D. discoideum* (Klein *et al., Science 241*:1467-1472 (1988)); and Family V, the fungal mating pheromone receptors such as STE2 (Kurjan, *Annu. Rev. Biochem. 61*:1097-1129 (1992)).

There are also a small number of other proteins which present seven putative hydrophobic segments and appear to be unrelated to GPCRs; they have not been shown to couple to G-proteins. *Drosophila* expresses a photoreceptor-specific protein, bride of sevenless (boss), a seven-transmembrane-segment protein which has been extensively studied and does not show evidence of being a GPCR (Hart *et al., Proc. Natl. Acad. Sci. USA 90*:5047-5051 (1993)). The gene frizzled (*fz*) in *Drosophila* is also thought to be a protein with seven transmembrane segments. Like boss, fz has not been shown to couple to G-proteins (Vinson *et al., Nature 338*:263-264 (1989)).

BNSDOCID: <WO___0149847A2_I_>

G proteins represent a family of heterotrimeric proteins composed of α, β and γ subunits, that bind guanine nucleotides. These proteins are usually linked to cell surface receptors, e.g., receptors containing seven transmembrane segments. Following ligand binding to the GPCR, a conformational change is transmitted to the G protein, which causes the α-subunit to exchange a bound GDP molecule for a GTP molecule and to dissociate from the βγ-subunits. The GTP-bound form of the α-subunit typically functions as an effector-modulating moiety, leading to the production of second messengers, such as cAMP (e.g., by activation of adenyl cyclase), diacylglycerol or inositol phosphates. Greater than 20 different types of α-subunits are known in humans. These subunits associate with a smaller pool of β and γ subunits. Examples of mammalian G proteins include Gi, Go, Gq, Gs and Gt. G proteins are described extensively in Lodish *et al., Molecular Cell Biology*, (Scientific American Books Inc., New York, N.Y., 1995), the contents of which are incorporated herein by reference. GPCRs, G proteins and G protein-linked effector and second messenger systems have been reviewed in *The G-Protein Linked Receptor Fact Book*, Watson *et al.*, eds., Academic Press (1994).

Purinoceptors

Purines, and especially adenosine and adenine nucleotides, have a broad range of pharmacological effects mediated through cell-surface receptors. For a general review, see "Adenosine and Adenine Nucleotides" in *The G-Protein Linked Receptor Facts Book*, Watson *et al.* (Eds.) Academic Press (1994), pp. 19-31.

Some effects of ATP include the regulation of smooth muscle activity, stimulation of the relaxation of intestinal smooth muscle and bladder contraction, stimulation of platelet activation by ADP when released from vascular endothelium, and excitatory effects in the central nervous system. Some effects of adenosine include vasodilation, bronchoconstriction, immunosuppression, inhibition of platelet aggregation, cardiac depression, stimulation of *nociceptive afferants*, inhibition of neurotransmitter release, pre- and postsynaptic depressant action, reducing motor activity, depressing respiration, inducing sleep, relieving anxiety, and inhibition of release of factors, such as hormones.

Distinct receptors exist for adenosine and adenine nucleotides. Clinical actions of such analogs as methylxanthines, for example, theophylline and caffeine, are thought to achieve their effects by antagonizing adenosine receptors. Adenosine has a low affinity for adenine nucleotide receptors, while adenine nucleotides have a low affinity for adenosine receptors.

There are four accepted subtypes of adenosine receptors, designated A_1, A_{2A}, A_{2B}, and A_3. In addition, an A_4 receptor has been proposed based on labeling by 2-phenylaminoadenosine (Cornfield *et al., Mol. Pharmacol.* 42:552-561 (1992)).

P_{2X} receptors are ATP-gated cation channels (See *Neuropharmacology 36* (1977)). The proposed topology for P_{2X} receptors is two transmembrane regions, a large extracellular loop, and intracellular N and C-termini.

Numerous cloned receptors designated P_{2Y} have been proposed to be members of the G-protein coupled family. UDP, UTP, ADP, and ATP have been identified as agonists. To date, P_{2Y1-7} have been characterized although it has been proposed that P_{2Y7} may be a leukotriene B4 receptor (Yokomizo *et al., Nature 387*:620-624 (1997)). It is widely accepted, however, that $P_{2Y\ 1,\ 2,\ 4,}$ and $_6$ are members of the G-protein coupled family of P_{2Y} receptors.

At least three P_2 purinoceptors from the hematopoietic cell line HEL have been identified by intracellular calcium mobilization and by photoaffinity labeling (Akbar *et al., J. Biochem. 271*:18363-18567 (1996)).

The A_1 adenosine receptor was designated in view of its ability to inhibit adenylcyclase. The receptors are distributed in many peripheral tissues such as heart, adipose, kidney, stomach and pancreas. They are also found in peripheral nerves, for example intestine and vas deferens. They are present in high levels in the central nervous system, including cerebral cortex, hippocampus, cerebellum, thalamus, and striatum, as well as in several cell lines. Agonists and antagonists can be found on page 22 of *The G-Protein Linked Receptor Facts Book* cited above, herein incorporated by reference. These receptors are reported to inhibit adenylcyclase and voltage-dependent calcium chanels and to activate potassium chanels through a pertussis-toxin-sensitive G-protein suggested to be of the G_i/G_0

BNSDOCID: <WO___0149847A2_I_>

class. A_1 receptors have also been reported to induce activation of phospholipase C and to potentiate the ability of other receptors to activate this pathway.

The A_{2A} adenosine receptor has been found in brain, such as striatum, olfactory tubercle and nucleus accumbens. In the periphery, A_2 receptors mediate vasodilation, immunosuppression, inhibition of platelet aggregation, and gluconeogenesis. Agonists and antagonists are found in *The G-Protein Linked Receptor Facts Book* cited above on page 25, herein incorporated by reference. This receptor mediates activation of adenylcyclase through G_8.

The A_{2B} receptor has been shown to be present in human brain and in rat intestine and urinary bladder. Agonists and antagonists are discussed on page 27 of *The G-Protein Linked Receptor Facts Book* cited above, herein incorporated by reference. This receptor mediates the stimulation of cAMP through G_8.

The A_3 adenosine receptor is expressed in testes, lung, kidney, heart, central nervous system, including cerebral cortex, striatum, and olfactory bulb. A discussion of agonists and antagonists can be found on page 28 of *The G-Protein Linked Receptor Facts Book* cited above, herein incorporated by reference. The receptor mediates the inhibition of adenylcyclase through a pertussis-toxin-sensitive G-protein, suggested to be of the G_i/G_0 class.

The P_{2Y} purinoceptor shows a similar affinity for ATP and ADP with a lower affinity for AMP. The receptor has been found in smooth muscle, for example, taeni caeci and in vascular tissue where it induces vasodilation through endothelium-dependent release of nitric oxide. It has also been shown in avian erythrocytes. Agonists and antagonists are discussed on page 30 of *The G-Protein Linked Receptor Facts Book* cited above, herein incorporated by reference. The receptor function through activation of phosphoinositide metabolism through a pertussis-toxin-insensitive G-protein, suggested to be of the G_i/G_0 class.

Receptor for Human C5a Anaphylatoxin

Chemotaxis of phagocytic cells is a key event in host defense and inflammatory responses. The C5a receptor mediates the pro-inflammatory and chemotaxis actions of the complement anaphylatoxin C5a. This receptor stimulates chemotaxis granule enzyme release, superoxide anion production, and

upregulates expression and activity of the adhesion molecule MAC-1 and of CR-1, and mediates a decrease in cell surface glycoprotein 100, MEL-14, in anaphylaxis and in septic shock. This receptor is a member of the rhodopsin superfamily of receptors. In contrast to other receptors of this family (adrenergic, serotoninergic, dopaminergic, FSH/LH, substance P and substance K), the C5a receptor functions in a concentration gradient of ligand and internalizes bound receptor during chemotaxis.

Accordingly, GPCRs, and especially complement receptors and purinoceptors, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown GPCRs. The present invention advances the state of the art by providing novel seven-transmembrane proteins/GPCRs, including a previously unidentified human seven-transmembrane protein/GPCR having homology to purinoceptors.

SUMMARY OF THE INVENTION

It is an object of the invention to identify novel seven-transmembrane proteins/GPCRs.

It is a further object of the invention to provide novel seven-transmembrane protein/GPCR polypeptides that are useful as reagents or targets in seven-transmembrane protein/receptor assays applicable to treatment and diagnosis of seven-transmembrane protein/GPCR-mediated disorders.

It is a further object of the invention to provide polynucleotides corresponding to the novel seven-transmembrane protein/GPCR receptor polypeptides that are useful as targets and reagents in seven-transmembrane protein/receptor assays applicable to treatment and diagnosis of seven-transmembrane protein/GPCR-mediated disorders and useful for producing novel seven-transmembrane protein/receptor polypeptides by recombinant methods.

A specific object of the invention is to identify compounds that act as agonists and antagonists and modulate the expression of the novel seven-transmembrane proteins/receptors.

BNSDOCID: <WO___0149847A2_I_>

A further specific object of the invention is to provide compounds that modulate expression of the seven-transmembrane proteins/receptors for treatment and diagnosis of seven-transmembrane protein/GPCR- related disorders.

The invention is thus based on the identification of novel seven-transmembrane proteins/GPCRs, designated 39404, 38911, and 26904. As discussed more fully below, 39404 contains sequence homology or motifs/signatures that classify this protein in the GPCR superfamily, as a member of the rhodopsin and metabotropic families of G-protein coupled receptors.

The invention provides isolated 39404 polypeptides including a polypeptide having the amino acid sequence shown in SEQ ID NO:1, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Patent Deposit No. PTA-1847 on May 9, 2000.

The invention provides isolated 38911 polypeptides including a polypeptide having the amino acid sequence shown in SEQ ID NO:3, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Patent Deposit No. PTA-1654 on April 6, 2000.

The invention provides isolated 26904 polypeptides including a polypeptide having the amino acid sequence shown in SEQ ID NO:5.

The invention also provides isolated 39404 nucleic acid molecules having the sequence shown in SEQ ID NO:2 or in the corresponding deposited cDNA.

The invention also provides isolated 38911 nucleic acid molecules having the sequence shown in SEQ ID NO:4 or in the corresponding deposited cDNA.

The invention also provides isolated 26904 nucleic acid molecules having the sequence shown in SEQ ID NO:6.

The invention also provides variant polypeptides having an amino acid sequence that is substantially homologous to the amino acid sequence shown in SEQ ID NO:1 or encoded by the deposited cDNA.

The invention also provides variant polypeptides having an amino acid sequence that is substantially homologous to the amino acid sequence shown in SEQ ID NO:3 or encoded by the deposited cDNA.

BNSDOCID: <WO___0149847A2_I_>

The invention also provides variant polypeptides having an amino acid sequence that is substantially homologous to the amino acid sequence shown in SEQ ID NO:5 or encoded by the deposited cDNA.

The invention also provides variant nucleic acid sequences that are substantially homologous to the nucleotide sequence shown in SEQ ID NO:2 or in the deposited cDNA.

The invention also provides variant nucleic acid sequences that are substantially homologous to the nucleotide sequence shown in SEQ ID NO:4 or in the deposited cDNA.

The invention also provides variant nucleic acid sequences that are substantially homologous to the nucleotide sequence shown in SEQ ID NO:6 or in the deposited cDNA.

The invention also provides fragments of the polypeptide shown in SEQ ID NO:1 and nucleotide sequence shown in SEQ ID NO:2, as well as substantially homologous fragments of the polypeptide or nucleic acid.

The invention also provides fragments of the polypeptide shown in SEQ ID NO:3 and nucleotide sequence shown in SEQ ID NO:4, as well as substantially homologous fragments of the polypeptide or nucleic acid.

The invention also provides fragments of the polypeptide shown in SEQ ID NO:5 and nucleotide sequence shown in SEQ ID NO:6, as well as substantially homologous fragments of the polypeptide or nucleic acid.

The invention further provides nucleic acid constructs comprising the nucleic acid molecules described above. In a preferred embodiment, the nucleic acid molecules of the invention are operatively linked to a regulatory sequence.

The invention also provides vectors and host cells for expressing the nucleic acid molecules and polypeptides of the invention and particularly recombinant vectors and host cells.

The invention also provides methods of making the vectors and host cells and methods for using them to produce the nucleic acid molecules and polypeptides of the invention.

The invention also provides antibodies or antigen-binding fragments thereof that selectively bind the polypeptides and fragments of the invention.

BNSDOCID: <WO 0149847A2 I_>

The invention also provides methods of screening for compounds that modulate expression or activity of the polypeptides or nucleic acid (RNA or DNA) of the invention.

The invention also provides a process for modulating polypeptide or nucleic acid expression or activity, especially using the screened compounds. Modulation may be used to treat conditions related to aberrant activity or expression of the polypeptides or nucleic acids of the invention.

The invention also provides assays for determining the presence or absence of and level of the polypeptides or nucleic acid molecules of the invention in a biological sample, including for disease diagnosis.

The invention also provides assays for determining the presence of a mutation in the polypeptides or nucleic acid molecules, including for disease diagnosis.

In still a further embodiment, the invention provides a computer readable means containing the nucleotide and/or amino acid sequences of the nucleic acids and polypeptides of the invention.

DESCRIPTION OF THE DRAWINGS

Figure 1 shows the 39404 nucleotide sequence (SEQ ID NO:2) and the deduced 39404 amino acid sequence (SEQ ID NO:1).

Figure 2 shows a 39404 protein hydrophobicity plot. The amino acids correspond to 1-337 and show the seven transmembrane segments.

Figure 3 shows an analysis of the 39404 amino acid sequence: αβturn and coil regions; hydrophilicity; amphipathic regions; flexible regions; antigenic index; and surface probability plot.

Figure 4 shows an analysis of the 39404 open reading frame for amino acids corresponding to specific functional sites. Glycosylation sites are shown in the figure with the actual modified residue being the first amino acid. cAMP- and cGMP-dependent protein kinase phosphorylation sites are shown in the figure with the actual modified residue being the last amino acid. Protein kinase C phosphorylation sites shown in the figure with the actual modified residue being

the first amino acid. A casein kinase II phosphorylation site is shown in the figure with the actual modified residue being the first amino acid. In addition, amino acids corresponding in position to the GPCR signature and containing the invariant arginine are found in the sequence FRY at amino acids 130-132. This figure also shows transmembrane segments predicted by MEMSAT for the predicted entire coding sequence and for a predicted mature peptide. For example, for the entire coding sequence, it is predicted that amino acids 1 to about 37 constitute the amino terminal extracellular domain, amino acids about 38-305 constitute the region spanning the transmembrane domain, and amino acids about 306-337 constitute the carboxy terminal intracellular domain. The transmembrane domain contains seven transmembrane segments, three extracellular loops and three intracellular loops. The transmembrane segments are found from about amino acid 38 to about amino acid 60, from about amino acid 70 to about amino acid 90, from about amino acid 117 to about amino acid 136, from about amino acid 149 to about amino acid 172, from about amino acid 200 to about amino acid 222, from about amino acid 242 to about amino acid 260, and from about amino acid 283 to about amino acid 305. Within the region spanning the entire transmembrane domain are three intracellular and three extracellular loops. The three intracellular loops are found from about amino acid 61 to about amino acid 69, from about amino acid 137 to about amino acid 148, and from about amino acid 223 to about amino acid 241. The three extracellular loops are found at from about amino acid 91 to about amino acid 116, from about amino acid 173 to about amino acid 199, and from about amino acid 261 to about amino acid 282.

Figure 5 shows expression of the 39404 protein in normal human tissues.

Figure 6 shows expression of the 39404 protein in human cardiovascular tissues. Int prolif: intimal proliferation; Int mamm: internal mammary; CHF: congestive heart failure; ISCH: ischemia; Myop: myopathy.

Figure 7 shows expression of the protein in human cardiovascular tissues.

Figure 8 shows the 38911 nucleotide sequence (SEQ ID NO:4) and the deduced 38911 amino acid sequence (SEQ ID NO:3).

BNSDOCID: <WO___0149847A2_I_>

Figure 9 shows an analysis of the 38911 amino acid sequence: αβturn and coil regions; hydrophilicity; amphipathic regions; flexible regions; antigenic index; and surface probability plot.

Figure 10 shows a 38911 protein hydrophobicity plot. The amino acids correspond to 1-337 and show the seven transmembrane segments.

Figure 11 shows an analysis of the 38911 open reading frame for amino acids corresponding to specific functional sites. A glycosylation site is found at amino acids 3-6. A cAMP- and cGMP-dependent protein kinase phosphorylation site is found at amino acids 324-327. A protein kinase C phosphorylation site is found at amino acids 17-19. A second protein kinase C phosphorylation site is found at amino acids 323-325. Casein kinase II phosphorylation sites are found at amino acids 194-197, 327-330, and 333-336. N-myristoylation sites are found at amino acids 26-31, 49-54, 103-108, 150-155, 156-161, 191-196, 253-258, 278-283, and 316-321. For the cAMP and cGMP dependent protein kinase phosphorylation, the actual modified residue is the last amino acid. For protein kinase C phosphorylation, the actual modified residue is the first amino acid. For casein kinase II phosphorylation, the actual modified residue is the first amino acid. For N-myristoylation, the actual modified residue is the first amino acid.

It is predicted that amino acids 1 to about 40 constitute the amino terminal extracellular domain, amino acids about 41-294 constitute the region spanning the transmembrane domain, and amino acids about 259-337 constitute the carboxy terminal intracellular domain. The transmembrane domain contains seven transmembrane segments, three extracellular loops and three intracellular loops. The transmembrane segments are found from about amino acid 41 to about amino acid 60, from about amino acid 68 to about amino acid 92, from about amino acid 113 to about amino acid 137, from about amino acid 153 to about amino acid 172, from about amino acid 205 to about amino acid 228, from about amino acid 237 to about amino acid 260, and from about amino acid 275 to about amino acid 294. Within the region spanning the entire transmembrane domain are three intracellular and three extracellular loops. The three intracellular loops are found from about amino acid 61 to about amino acid 67, from about amino acid 138 to about amino acid 152, and from about amino acid 229 to about amino acid 236.

The three extracellular loops are found at from about amino acid 93 to about amino acid 112, from about amino acid 173 to about amino acid 204, and from about amino acid 261 to about amino acid 274.

Figure 12 shows expression of the 38911 protein in various normal human tissues, using fetal heart as a reference.

Figure 13 shows expression of the 38911 protein in various normal human tissues and in biopsies from fibrotic livers.

Figure 14 shows the 26904 nucleotide sequence (SEQ ID NO:6) and the deduced 26904 amino acid sequence (SEQ ID NO:5).

Figure 15 shows an analysis of the 26904 amino acid sequence: αβturn and coil regions; hydrophilicity; amphipathic regions; flexible regions; antigenic index; and surface probability plot.

Figure 16 shows a 26904 protein hydrophobicity plot. The amino acids show the seven transmembrane segments.

Figure 17 shows an analysis of the 26904 open reading frame for amino acids corresponding to specific functional sites. A glycosylation site is found at amino acids 312-315. A cAMP- and cGMP-dependent protein kinase phosphorylation site is found at amino acids 143-146. Protein kinase C phosphorylation sites are found at about amino acids 6-8, 136-138, 234-236, 245-247, 314-316, 436-438, and 446-448. Casein kinase II phosphorylation sites are found at about amino acids 55-58, 167-170, 218-221, 239-242, 284-287, 416-419, and 447-450. Tyrosine kinase phosphorylation sites are found at about amino acids 118-125, 336-343, 382-389, and 409-415. N-myristoylation sites are found at about amino acids 36-41, 91-96, 261-266, 304-309, 365-370, 404-409, and 420-425. An amidation site is found at about amino acids 141-144. An ATP/GTP-binding site motif A (P-loop) is found at about amino acids 230-237. In the case of protein kinase C phosphorylation, the actual modified residue is the first amino acid. In the case of casein kinase II phosphorylation, the actual modified residue is the first amino acid. In the case of the tyrosine kinase phosphorylation, the modified amino acid is the last amino acid. In the case of N-myristoylation, the modified amino acid is the first amino acid.

ENSDOCID <WO 0149847A2 I >

It is predicted that amino acids 1 to about 30 constitute the amino terminal extracellular domain, amino acids about 30-435 constitute the region spanning the transmembrane domain, and amino acids about 435-450 constitute the carboxy terminal intracellular domain. The transmembrane domain contains seven transmembrane segments, three extracellular loops and three intracellular loops. The transmembrane segments are found from about amino acid 30 to about amino acid 50, from about amino acid 100 to about amino acid 120, from about amino acid 140 to about amino acid 165, from about amino acid 200 to about amino acid 240, from about amino acid 305 to about amino acid 340, from about amino acid 360 to about amino acid 380, and from about amino acid 410 to about amino acid 450. Within this region spanning the entire transmembrane domain are three intracellular and three extracellular loops.

Figure 18 shows the expression of 38911 in the following tissues: normal human lung (column 1), normal human kidney (column 2), normal human brain (column 3), normal human granulocytes (column 4), normal human heart (column 5), normal human spleen (column 6), normal human fetal liver (column 7), a pool of 7 normal human livers (column 8), resting normal human dermal fibroblasts (column 9), normal human lung fibroblasts (column 10), normal human lung fibroblasts cultured for 48 hours with TGF-β (column 11), human fibrotic liver (columns 12-15), normal human tonsils (column 16), proinflammatory type 1T helper cells (column 17 and 19) and proinflammatory type 2T cells (column 18 and 20). 38911 was expressed at high levels in kidney, spleen, fetal liver, fibrotic liver, and tonsils, and at moderate levels in lung, brain, granulocytes, heart, and normal liver. Expression levels were determined by quantitative PCR (Taqman® brand quantitative PCR kit, Applied Biosystems). The quantitative PCR reactions were performed according to the kit manufacturer's instructions.

Figure 19 shows the expression of 38911 in the following tissues: CD4+ cells (column 1), CD8+cells (column 2), resting CD14+ monocytes (column 3), resting peripheral blood mononuclear cells (column 4), CD19+ cells (column 5), resting CD3+ cells (column 6), bone marrow mononuclear cells (column 7), mobilized peripheral blood CD34+ cells (column 8), adult bone marrow CD34+ cells (column 9), human cord blood CD34+ cells (column 10), human erythroid

cells (column11), human megakaryocytes (column 12), cultured day 14 neutrophiles (column 13), mobilized bone marrow CD15+ cells (column 14), GPA+ cells from human bone marrow (column 15), HepG2.2 cells transfected with hepatitis B virus (column 16), hepatitis B virus-infected liver (column 17), hepatoma Hep3B cells cultured with normal oxygen level (column 18), hepatoma Hep3B cells cultured with low levels of oxygen (column 19). 38911 expression levels were determined as described in the figure legend for Figure 18.

DETAILED DESCRIPTION OF THE INVENTION

Receptor function/signal pathway

The 39404 38911, and 26904 receptor proteins are GPCR-like proteins that participate in signaling pathways. As used herein, a "signaling pathway" refers to the modulation (e.g., stimulation or inhibition) of a cellular function/activity upon the binding of a ligand to the GPCR (39404, 38911, or 26904 protein). Examples of such functions include mobilization of intracellular molecules that participate in a signal transduction pathway, e.g., phosphatidylinositol 4,5-bisphosphate (PIP_2), inositol 1,4,5-triphosphate (IP_3) and adenylate cyclase; polarization of the plasma membrane; production or secretion of molecules; alteration in the structure of a cellular component; cell proliferation, e.g., synthesis of DNA; cell migration; cell differentiation; and cell survival. The 39404 protein is expressed in the tissues shown in Figures 5-7. Therefore, cells participating in a 39404 protein signaling pathway include, but are not limited to, cells derived from these tissues, especially those tissues in which the gene is highly expressed, such as brain, kidney, aortic intimal proliferations, and internal mammary artery. Since the 38911 protein is expressed in the tissues shown in Figures 12 and 13, cells participating in a 38911 protein signaling pathway include, but are not limited to, cells derived from these tissues, especially those cells or tissues in which the gene is highly expressed, such as osteoclasts, spleen, liver, kidney, tonsils, and testis. The gene is also expressed in $CD4^+$ cells (T-lymphocytes), in peripheral blood monocytes, and in neutrophils. Since the 26904 protein is expressed in brain, cells participating in a 26904 protein signaling pathway include, but are not limited to, cells derived from this tissue.

BNSDOCID: <WO___0149847A2_I_>

The response mediated by a receptor protein depends on the type of cell. For example, in some cells, binding of a ligand to the receptor protein may stimulate an activity such as release of compounds, gating of a channel, cellular adhesion, migration, differentiation, etc., through phosphatidylinositol or cyclic AMP metabolism and turnover while in other cells, the binding of the ligand will produce a different result. Regardless of the cellular activity/response modulated by the receptor protein, it is universal that a GPCR of the invention interacts with G proteins to produce one or more secondary signals, in a variety of intracellular signal transduction pathways, e.g., through phosphatidylinositol or cyclic AMP metabolism and turnover, in a cell.

As used herein, "phosphatidylinositol turnover and metabolism" refers to the molecules involved in the turnover and metabolism of phosphatidylinositol 4,5-bisphosphate (PIP_2) as well as to the activities of these molecules. PIP_2 is a phospholipid found in the cytosolic leaflet of the plasma membrane. Binding of ligand to the receptor activates, in some cells, the plasma-membrane enzyme phospholipase C that in turn can hydrolyze PIP_2 to produce 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP_3). Once formed IP_3 can diffuse to the endoplasmic reticulum surface where it can bind an IP_3 receptor, e.g., a calcium channel protein containing an IP_3 binding site. IP_3 binding can induce opening of the channel, allowing calcium ions to be released into the cytoplasm. IP_3 can also be phosphorylated by a specific kinase to form inositol 1,3,4,5-tetraphosphate (IP_4), a molecule which can cause calcium entry into the cytoplasm from the extracellular medium. IP_3 and IP_4 can subsequently be hydrolyzed very rapidly to the inactive products inositol 1,4-biphosphate (IP_2) and inositol 1,3,4-triphosphate, respectively. These inactive products can be recycled by the cell to synthesize PIP_2. The other second messenger produced by the hydrolysis of PIP_2, namely 1,2-diacylglycerol (DAG), remains in the cell membrane where it can serve to activate the enzyme protein kinase C. Protein kinase C is usually found soluble in the cytoplasm of the cell, but upon an increase in the intracellular calcium concentration, this enzyme can move to the plasma membrane where it can be activated by DAG. The activation of protein kinase C in different cells results in various cellular responses such as the phosphorylation of glycogen synthase, or the

phosphorylation of various transcription factors, e.g., NF-kB. The language "phosphatidylinositol activity", as used herein, refers to an activity of PIP_2 or one of its metabolites.

Another signaling pathway in which a receptor protein of the invention may participate is the cAMP turnover pathway. As used herein, "cyclic AMP turnover and metabolism" refers to the molecules involved in the turnover and metabolism of cyclic AMP (cAMP) as well as to the activities of these molecules. Cyclic AMP is a second messenger produced in response to ligand-induced stimulation of certain G protein coupled receptors. In the cAMP signaling pathway, binding of a ligand to a GPCR can lead to the activation of the enzyme adenyl cyclase, which catalyzes the synthesis of cAMP. The newly synthesized cAMP can in turn activate a cAMP-dependent protein kinase. This activated kinase can phosphorylate a voltage-gated potassium channel protein, or an associated protein, and lead to the inability of the potassium channel to open during an action potential. The inability of the potassium channel to open results in a decrease in the outward flow of potassium, which normally repolarizes the membrane of a neuron, leading to prolonged membrane depolarization.

Polypeptides

The invention is based on the identification of novel seven-transmembrane proteins/G-coupled protein receptors. Specifically, an expressed sequence tag (EST) was selected based on homology to G-protein-coupled receptor sequences or motifs (e.g.., seven-transmembrane domains). This EST was used to design primers based on sequences that it contains and used to identify a 39404 cDNA from a human colon cDNA library, a 38911 cDNA from a human bone marrow cDNA library, and a 26904 cDNA from a human brain cDNA library. Positive clones were sequenced and the overlapping fragments were assembled. Analysis of the assembled sequences revealed that the cloned cDNA molecules encode G-protein coupled receptors (39404, 38911, 26904).

The invention thus relates to a novel GPCR having the deduced amino acid sequence shown in Figure 1 (SEQ ID NO:1) or having the amino acid sequence

BNSDOCID: <WO___0149847A2_I_>

encoded by the cDNA insert of the plasmid deposited with ATCC as Patent Deposit No. PTA-1847.

The invention also thus relates to a novel putative GPCR having the deduced amino acid sequence shown in Figure 8 (SEQ ID NO:3) or having the amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Patent Deposit No. PTA-1654.

The invention also thus relates to a novel putative GPCR having the deduced amino acid sequence shown in Figure 14 (SEQ ID NO:5).

Plasmids containing the nucleotide sequences of the invention were deposited with the Patent Depository of the American Type Culture Collection (ATCC), Manassas, Virginia. The deposits will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms. The deposits are provided as a convenience to those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112. The deposited sequences, as well as the polypeptides encoded by the sequences, are incorporated herein by reference and control in the event of any conflict, such as a sequencing error, with description in this application.

The "39404 polypeptide" or "39404 protein" refers to the polypeptide in *SEQ ID NO:1 or encoded by the deposited cDNA. The "38911 polypeptide" or* "38911 protein" refers to the polypeptide in SEQ ID NO:3 or encoded by the deposited cDNA. The "26904 polypeptide" or "26904 protein" refers to the polypeptide in SEQ ID NO:5 or encoded by the deposited cDNA. The term "protein" or "polypeptide", however, further includes the numerous variants of 39404, 38911, or 26904 polypeptides described herein, as well as fragments derived from the full length 39404, 38911, or 26904 polypeptides and variants.

The present invention thus provides isolated or purified 39404, 38911, and 26904 polypeptides and variants and fragments thereof.

The 39404 polypeptide is a 319 residue protein exhibiting three main *structural domains, an amino terminal extracellular domain, transmembrane* domain, and carboxy terminal intracellular domain, as shown in Figure 4. Based on a BLAST search, highest homology was shown to purinoceptors (rhodopsin superfamily)

The 38911 polypeptide is a 337 residue protein exhibiting three main structural domains, the amino terminal extracellular domain, transmembrane domain, and carboxy terminal intracellular domain, as shown in Figure 11. Based on a BLAST search, highest homology was shown to the C5a anaphylatoxin receptor (G-protein Linked Receptor Facts Book , Watson and Arkinstall, Editors, Academic Press (1994) New York, pgs. 71-73, incorporated herein by reference for its teachings regarding this receptor).

The 26904 polypeptide is a 450 residue protein exhibiting three main structural domains, the amino terminal extracellular domain, transmembrane domain, and carboxy terminal intracellular domain, as shown in Figure 11.

As used herein, a polypeptide is said to be "isolated" or "purified" when it is substantially free of cellular material when it is isolated from recombinant and non-recombinant cells, or free of chemical precursors or other chemicals when it is chemically synthesized. A polypeptide, however, can be joined to another polypeptide with which it is not normally associated in a cell and still be considered "isolated" or "purified."

The polypeptides of the invention can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to homogeneity are useful and considered to contain an isolated form of the polypeptide. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity.

In one embodiment, the language "substantially free of cellular material" includes preparations of the polypeptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the polypeptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the protein preparation.

A polypeptide is also considered to be isolated when it is part of a membrane preparation or is purified and then reconstituted with membrane vesicles or liposomes.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

In one embodiment, the 39404 polypeptide comprises the amino acid sequence shown in SEQ ID NO:1. However, the invention also encompasses sequence variants. Variants include a substantially homologous protein encoded by the same genetic locus in an organism, i.e., an allelic variant. Variants also encompass proteins derived from other genetic loci in an organism, but having substantial homology to the 39404 protein of SEQ ID NO:1. Variants also include proteins substantially homologous to the 39404 protein but derived from another organism, i.e., an ortholog. Variants also include proteins that are substantially homologous to the 39404 protein that are produced by chemical synthesis. Variants also include proteins that are substantially homologous to the 39404 protein that are produced by recombinant methods. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

In another embodiment, the 38911 polypeptide comprises the amino acid sequence shown in SEQ ID NO:3. However, the invention also encompasses sequence variants. Variants include allelic variants. Variants also encompass proteins derived from other genetic loci in an organism, but having substantial homology to the 38911 protein of SEQ ID NO:3. Variants also include substantially homologous orthologs. Variants also include proteins that are substantially homologous to the 38911 protein that are produced by chemical synthesis. Variants also include proteins that are substantially homologous to the 38911 protein that are produced by recombinant methods. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

In another embodiment, the 26904 polypeptide comprises the amino acid sequence shown in SEQ ID NO:5. However, the invention also encompasses sequence variants. Variants include allelic variants. Variants also encompass proteins derived from other genetic loci in an organism, but having substantial homology to the 26904 protein of SEQ ID NO:5. Variants also include substantially homologous orthologs. Variants also include proteins that are substantially homologous to the 26904 protein that are produced by chemical synthesis. Variants also include proteins that are substantially homologous to the 26904 protein that are produced by recombinant methods. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

As used herein, two proteins (or a region of the proteins) are substantially homologous to the 39404 protein when the amino acid sequences are at least about 40-45%, 45-50%, 50-55%, 55-60%, typically at least about 60-65%, 65-70%, or 70-75%, more typically at least about 70-75%, 75-80%, or 80-85%, and most typically at least about 85-90% or 90-95% or more homologous. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence hybridizing to the nucleic acid sequence, or portion thereof, of the sequence shown in SEQ ID NO:2 under stringent conditions as more fully described below.

As used herein, two proteins (or a region of the proteins) are substantially homologous to the 38911 protein when the amino acid sequences are at least about 35-40%, 40-45%, 45-50%, 55-60%, 60-65%, 65-70%, typically at least about 70-75%, more typically at least about 75-80% or 80-85%, and most typically at least about 85-90% or 90-95% or more homologous. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence hybridizing to the nucleic acid sequence, or portion thereof, of the sequence shown in SEQ ID NO:4 under stringent conditions as more fully described below.

As used herein, two proteins (or a region of the proteins) are substantially homologous to the 26904 protein when the amino acid sequences are at least about 50-55%, 55-60%, 60-65%, 65-70%, typically at least about 70-75%, more

BNSDOCID: <WO___0149847A2_I_>

typically at least about 75-80% or 80-85%, and most typically at least about 85-90% or 90-95% or more homologous. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence hybridizing to the nucleic acid sequence, or portion thereof, of the sequence shown in SEQ ID NO:6 under stringent conditions as more fully described below.

The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions (*e.g.* G protein coupled signaling) performed by the 39404, 38911, or 26904 polypeptides. Similarity is determined by conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science 247*:1306-1310 (1990).

TABLE 1. Conservative Amino Acid Substitutions.

Aromatic	Phenylalanine
	Tryptophan
	Tyrosine
Hydrophobic	Leucine
	Isoleucine
	Valine
Polar	Glutamine
	Asparagine
Basic	Arginine
	Lysine
	Histidine
Acidic	Aspartic Acid
	Glutamic Acid
Small	Alanine
	Serine
	Threonine
	Methionine
	Glycine

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of

BNSDOCID: <WO___0149847A2_I_>

the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished by well-known methods such as using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., Ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., Ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part 1, Griffin, A.M., and Griffin, H.G., Eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., Eds., M Stockton Press, New York, 1991).

A preferred, non-limiting example of such a mathematical algorithm is described in Karlin *et al.* (1993) *Proc. Natl. Acad. Sci.* USA *90*:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) as described in Altschul *et al.* (1997) *Nucleic Acids Res. 25*:3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., NBLAST) can be used. See www.ncbi.nlm.nih.gov. In one embodiment, parameters for sequence comparison can be set at score= 100, wordlength = 12, or can be varied (e.g., W=5 or W=20).

In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman *et al.* (1970) (*J. Mol. Biol. 48*:444-453) algorithm which has been incorporated into the GAP program in the GCG software package (available at www.gcg.com), using either a BLOSUM 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length

BNSDOCID: <WO___0149847A2_I_>

weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux *et al.* (1984) *Nucleic Acids Res. 12*(1):387) (available at www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6.

Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the CGC sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12 , and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art and include ADVANCE and ADAM as described in Torellis *et al* (1994) *Comput. Appl. Biosci. 10*:3-5; and FASTA described in Pearson *et al.* (1988) *PNAS 85*:2444-8.

A variant polypeptide can differ in amino acid sequence by one or more substitutions, deletions, insertions, inversions, fusions, and truncations or a combination of any of these.

Variant polypeptides can be fully functional or can lack function in one or more activities. Thus, in the present case, variations can affect the function, for example, of one or more of the regions corresponding to ligand binding, membrane association, G-protein binding and signal transduction.

Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids which result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

As indicated, variants can be naturally-occurring or can be made by recombinant means or chemical synthesis to provide useful and novel

BNSDOCID: <WO___0149847A2_I_>

characteristics for the polypeptide. This includes preventing immunogenicity from pharmaceutical formulations by preventing protein aggregation.

Useful variations further can include alteration of ligand binding characteristics. For example, one embodiment involves a variation at the binding site that results in binding but not release, or slower release, of ligand. A further useful variation at the same sites can result in a higher affinity for ligand. Useful variations also include changes that provide for affinity for another ligand. Another useful variation can include one that allows binding but which prevents activation by the ligand. Another useful variation includes variation in the transmembrane G-protein-binding/signal transduction domain that provides for reduced or increased binding by the appropriate G-protein or for binding by a different G-protein than the one with which the receptor is normally associated. Another useful variation provides a fusion protein in which one or more domains or subregions is operationally fused to one or more domains or subregions from another G-protein coupled receptor.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al., Science 244*:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro*, or *in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al., J. Mol. Biol. 224*:899-904 (1992); de Vos *et al. Science 255*:306-312 (1992)).

Substantial homology can be to the entire nucleic acid or amino acid sequence or to fragments of these sequences.

The invention thus also includes polypeptide fragments of the 39404 protein. Fragments can be derived from the amino acid sequence shown in SEQ ID NO:1. However, the invention also encompasses fragments of the variants of the 39404 protein as described herein.

The invention thus also includes polypeptide fragments of the 38911 protein. Fragments can be derived from the amino acid sequence shown in SEQ

ID NO:3. However, the invention also encompasses fragments of the variants of the 38911 protein as described herein.

The invention thus also includes polypeptide fragments of the 26904 protein. Fragments can be derived from the amino acid sequence shown in SEQ ID NO:5. However, the invention also encompasses fragments of the variants of the 26904 protein as described herein.

The fragments *per se* to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed prior to the present invention (known fragments are encompassed in uses and methods specific for tissues or disorders with which the gene is associated).

Fragments can retain one or more of the biological activities of the protein, for example, the ability to bind to a G-protein or ligand, as well as fragments that can be used as an immunogen to generate antibodies.

Biologically active fragments of the 39404 protein (peptides which are, for example, 5-10, 10-15, 15-20, 20-30, 30-40, 40-50, 50-100, or more amino acids in length) can comprise a domain or motif, e.g., an extracellular or intracellular domain or loop, one or more transmembrane segments or parts thereof, G-protein binding site, GPCR signature, glycosylation site, or phosphorylation site. In one embodiment, fragments are greater than eleven amino acids.

Such domains or motifs can be identified by means of routine computerized homology searching procedures.

Possible fragments include, but are not limited to: 1) soluble peptides comprising the entire amino terminal extracellular domain or parts thereof; 2) peptides comprising the entire carboxy terminal intracellular domain or parts thereof; 3) peptides comprising the region spanning the entire transmembrane domain or parts thereof; 4) any of the specific transmembrane segments, or parts thereof; 5) any of the three intracellular or three extracellular loops, or parts thereof. Fragments further include combinations of the above fragments, such as an amino terminal domain combined with one or more transmembrane segments and the attendant extra or intracellular loops or one or more transmembrane segments, and the attendant intra or extracellular loops, plus the carboxy terminal domain. Thus, any of the above fragments can be combined. Other fragments

BNSDOCID: <WO__0149847A2_I_>

include the mature protein from about amino acid 6 to the last amino acid. Other fragments contain the various functional sites described herein, such as phosphorylation sites or glycosylation sites, and a sequence containing the GPCR signature sequence. Fragments, for example, can extend in one or both directions from the functional site to encompass 5, 10, 15, 20, 30, 40, 50, or up to 100 amino acids. Further, fragments can include sub-fragments of the specific domains mentioned above, which sub-fragments retain the function of the domain from which they are derived. Fragments also include but are not limited to amino acid sequences greater than 5 amino acids, except for SILTLT (SEQ ID NO:7), SILFLTC (SEQ ID NO:8), or NLYSSILFLTC (SEQ ID NO:9) (however, it is understood that with regard to uses and methods of the invention, even these fragments and any other fragments that may be known prior to the invention are encompassed). In no way however are such fragments to be construed as encompassing fragments that may be found in the art, except as above indicated.

These regions can be identified by well-known methods involving computerized homology analysis.

Fragments also include antigenic fragments and specifically in regions shown to have a high antigenic index in Figure 3.

Accordingly, possible fragments include fragments defining a ligand-binding site, fragments defining a glycosylation site, fragments defining membrane association, fragments defining phosphorylation sites, and fragments defining interaction with G proteins and signal transduction. By this is intended a discrete fragment that provides the relevant function or allows the relevant function to be identified. In a preferred embodiment, the fragment contains the ligand-binding site.

Biologically active fragments of 38911 protein (peptides which are, for example, 5-10, 10-15, 15-20, 20-30, 30-40, 40-50, 50-100, or more amino acids in length) can comprise a domain or motif, e.g., an extracellular or intracellular domain or loop, one or more transmembrane segments, or parts thereof, G-protein binding site, glycosylation sites, and cAMP- and cGMP-dependent, protein kinase C, and casein kinase II phosphorylation sites, and N-myristoylation sites.

Such domains or motifs can be identified by means of routine computerized homology searching procedures.

Possible fragments include, but are not limited to: 1) soluble peptides comprising the entire amino terminal extracellular domain about amino acid 1 to about amino acid 40 of SEQ ID NO:3, or parts thereof; 2) peptides comprising the entire carboxy terminal intracellular domain from about amino acid 259 to amino acid 337 of SEQ ID NO:3, or parts thereof; 3) peptides comprising the region spanning the entire transmembrane domain from about amino acid 41 to about amino acid 294, or parts thereof; 4) any of the specific transmembrane segments, or parts thereof; 5) any of the three intracellular or three extracellular loops, or parts thereof. Fragments further include combinations of the above fragments, such as an amino terminal domain combined with one or more transmembrane segments and the attendant extra or intracellular loops or one or more transmembrane segments, and the attendant intra or extracellular loops, plus the carboxy terminal domain. Thus, any of the above fragments can be combined. Other fragments include the mature protein from about amino acid 6 to 337. Other fragments contain the various functional sites described herein, such as phosphorylation sites, glycosylation sites, or myristoylation sites. Fragments, for example, can extend in one or both directions from the functional site to encompass 5, 10, 15, 20, 30, 40, 50, or up to 100 amino acids. Further, fragments can include sub-fragments of the specific domains mentioned above, which sub-fragments retain the function of the domain from which they are derived.

These regions can be identified by well-known methods involving computerized homology analysis.

Fragments also include amino acid sequences greater than 5 amino acids except for LAVADLL (SEQ ID NO:10), LALLLT (SEQ ID NO:11), LRRSLP (SEQ ID NO:12), FLVGDPGNA (SEQ ID NO:13), GNAMV (SEQ ID NO:14), LAVAD (SEQ ID NO:15), FLVGVPGNA (SEQ ID NO:16), ALLLT (SEQ ID NO:17), and ADLLCCLSLP (SEQ ID NO:18) (it is understood however that these fragments and any others that may have been disclosed prior to the invention may be encompassed in specific uses and methods disclosed herein relating to tissues/disorders with which the expression is associated). In no way however are

BNSDOCID: <WO 0149847A2 I >

such fragments to be construed as encompassing fragments that may be found in the art except as just indicated.

Fragments also include antigenic fragments and specifically from regions shown to have a high antigenic index in Figure 9.

Accordingly, possible fragments include fragments defining a ligand-binding site, fragments defining a glycosylation site, fragments defining membrane association, fragments defining a phosphorylation site, fragments defining interaction with G proteins and signal transduction, and fragments defining a myristoylation site. By this is intended a discrete fragment that provides the relevant function or allows the relevant function to be identified. In a preferred embodiment, the fragment contains the ligand-binding site.

Biologically active fragments of the 26904 protein (peptides which are, for example, 5-10, 10-15, 15-20, 20-30, 30-40, 40-50, 50-100, or more amino acids in length) can comprise a domain or motif, e.g., an extracellular or intracellular domain or loop, one or more transmembrane segments, or parts thereof, G-protein binding site, glycosylation site, cAMP, cGMP, protein kinase C, and casein kinase II phosphorylation site, N-myristoylation site, amidation, or ATP/GTP binding site.

Such domains or motifs can be identified by means of routine computerized homology searching procedures.

Possible fragments include, but are not limited to: 1) soluble peptides comprising the entire amino terminal extracellular domain, or parts thereof; 2) peptides comprising the entire carboxy terminal intracellular domain, or parts thereof; 3) peptides comprising the region spanning the entire transmembrane domain, or parts thereof; 4) any of the specific transmembrane segments, or parts thereof; 5) any of the three intracellular or three extracellular loops, or parts thereof. Fragments further include combinations of the above fragments, such as an amino terminal domain combined with one or more transmembrane segments and the attendant extra or intracellular loops or one or more transmembrane segments, and the attendant intra or extracellular loops, plus the carboxy terminal domain. Thus, any of the above fragments can be combined. Other fragments include the mature protein from about amino acid 6 to 450. Other fragments

contain the various functional sites described herein, such as phosphorylation sites, glycosylation site, or myristoylation sites. Fragments, for example, can extend in one or both directions from the functional site to encompass 5, 10, 15, 20, 30, 40, 50, or up to 100 amino acids. Further, fragments can include sub-fragments of the specific domains mentioned above, which sub-fragments retain the function of the domain from which they are derived.

These regions can be identified by well-known methods involving computerized homology analysis.

Fragments also include amino acid sequences greater than four amino acids except for YVGAAHG (SEQ ID NO:19), LVHWCHGAPGVI (SEQ ID NO:20), QAYKVF (SEQ ID NO:21), EEKYL (SEQ ID NO:22), SLFEGMAG (SEQ ID NO:23), RFPAFEL (SEQ ID NO:24), LLQQME (SEQ ID NO:25), TFLCGDAGPLAV (SEQ ID NO:26), AGIYY (SEQ ID NO:27), SGNYP (SEQ ID NO:28), QAYKVFKEE (SEQ ID NO:29), DVIWQ (SEQ ID NO:30), KYLYRACKFAEWCLDYG (SEQ ID NO:31), ELLYGR (SEQ ID NO:32), PYSLFEG (SEQ ID NO:33), and VTFLCG (SEQ ID NO:34) (it is understood however that these fragments and any others that may have been disclosed prior to the invention are in fact encompassed by the invention in methods and uses disclosed herein relevant to specific tissues or disorders with which the gene is associated). In no way however are such fragments to be construed as encompassing fragments that may be found in the art, except as just indicated.

Fragments also include antigenic fragments and specifically from sites shown to have a high antigenic index in Figure 15.

Accordingly, possible fragments include but are not limited to fragments defining a ligand-binding site, fragments defining a glycosylation site, fragments defining membrane association, fragments defining phosphorylation sites, fragments defining interaction with G proteins and signal transduction, and fragments defining myristoylation sites. By this is intended a discrete fragment that provides the relevant function or allows the relevant function to be identified. In a preferred embodiment, the fragment contains the ligand-binding site.

The invention also provides 39404 protein fragments with immunogenic properties. These contain an epitope-bearing portion of the 39404 protein and

ENSDOCID: <WO 0149847A2 I >

variants. The invention also provides 38911 protein fragments with immunogenic properties. These contain an epitope-bearing portion of the 38911 protein and variants. The invention also provides 26904 protein fragments with immunogenic properties. These contain an epitope-bearing portion of the 26904 protein and variants. These peptides can contain at least 5-10, 11, 12, 13, at least 14, or between at least about 15 to about 30 amino acids.

Non-limiting examples of antigenic polypeptides that can be used to generate antibodies include peptides derived from the amino terminal extracellular domain or any of the extracellular loops. Regions having a high antigenicity index are shown in Figures 3, 9, and 15.

The epitope-bearing receptor and polypeptides may be produced by any conventional means (Houghten, R.A., *Proc. Natl. Acad. Sci. USA 82*:5131-5135 (1985)). Simultaneous multiple peptide synthesis is described in U.S. Patent No. 4,631,211.

Fragments can be discrete (not fused to other amino acids or polypeptides) or can be within a larger polypeptide. Further, several fragments can be comprised within a single larger polypeptide. In one embodiment a fragment designed for expression in a host can have heterologous pre- and pro-polypeptide regions fused to the amino terminus of the fragment and an additional region fused to the carboxyl terminus of the fragment.

The invention thus provides chimeric or fusion proteins. These comprise a protein of the invention operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the protein. "Operatively linked" indicates that the protein of the invention and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the protein of the invention.

In one embodiment the fusion protein does not affect protein function *per se*. For example, the fusion protein can be a GST-fusion protein in which the sequences of the invention are fused to the C-terminus of the GST sequences. Other types of fusion proteins include, but are not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL-4 fusions, poly-His fusions and Ig fusions. Such fusion proteins, particularly poly-His

fusions, can facilitate the purification of recombinant protein of the invention. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence. Therefore, in another embodiment, the fusion protein contains a heterologous signal sequence at its C- or N-terminus.

EP-A-O 464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists. Bennett *et al.* (*J. Mol. Recog. 8*:52-58 (1995)) and Johanson *et al. (J. Biol. Chem. 270, 16*:9459-9471 (1995)). Thus, this invention also encompasses soluble fusion proteins containing a polypeptide of the invention and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclass (IgG, IgM, IgA, IgE). Preferred as immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. For some uses it is desirable to remove the Fc after the fusion protein has been used for its intended purpose, for example when the fusion protein is to be used as antigen for immunizations. In a particular embodiment, the Fc part can be removed in a simple way by a cleavage sequence which is also incorporated and can be cleaved with factor Xa.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology,* 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A protein-encoding

BNSDOCID: <WO___0149847A2_I_>

nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein.

Another form of fusion protein is one that directly affects protein functions. Accordingly, a polypeptide is encompassed by the present invention in which one or more of the receptor domains (or parts thereof) has been replaced by homologous domains (or parts thereof) from another seven-transmembrane protein, for example another G-protein coupled receptor or other type of receptor. Accordingly, various permutations are possible. The amino terminal extracellular domain, or subregion thereof, (for example, ligand-binding) can be replaced with the domain or subregion from another ligand-binding receptor protein. Alternatively, the entire transmembrane domain, or any of the seven segments or loops, or parts thereof, for example, G-protein-binding/signal transduction, can be replaced. Finally, the carboxy terminal intracellular domain or subregion can be replaced. Thus, chimeric seven-transmembrane proteins/receptors can be formed in which one or more of the native domains or subregions has been replaced.

The isolated 39404 protein can be purified from cells that naturally express it, such as from breast, brain, kidney, vein, fetal kidney and fetal liver, shown in Figure 5, as well as aortic intimal proliferations and internal mammary artery as shown in Figures 6 and 7, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods.

The isolated 38911 protein can be purified from cells that naturally express it, such as from those tissues shown in Figures 12 and 13, and especially osteoclasts, spleen, tonsils, liver, kidney, and testis, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods.

The isolated 26904 protein can be purified from cells that naturally express it, such as from brain, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods.

In one embodiment, the protein is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from

the cells by an appropriate purification scheme using standard protein purification techniques.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally-occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in polypeptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art.

Accordingly, the polypeptides also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature polypeptide or a pro-protein sequence.

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of *phosphotidylinositol, cross-linking, cyclization, disulfide bond formation,* demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for

BNSDOCID: <WO__0149847A2_I_>

instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol. 182*: 626-646 (1990)) and Rattan *et al. Ann. N.Y. Acad. Sci. 663*:48-62 (1992)).

As is also well known, polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of post-translation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translational natural processes and by synthetic methods.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. Blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally-occurring and synthetic polypeptides. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

The modifications can be a function of how the protein is made. For recombinant polypeptides, for example, the modifications will be determined by the host cell posttranslational modification capacity and the modification signals in the polypeptide amino acid sequence. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to efficiently express mammalian proteins having native patterns of glycosylation. Similar considerations apply to other modifications.

The same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain more than one type of modification.

Polypeptide uses

The polypeptides are useful for various biological assays as described in detail below. Since the 39404 gene is expressed in the tissues shown in Figures 5-7, the assays are particularly useful in cells derived from these tissue types, and particularly the tissues in which the gene is highly expressed, such as brain, kidney, fetal kidney, fetal liver, internal mammary artery, and aortic intimal proliferations. Furthermore, since the gene is expressed in these tissues, assays involving the protein in pathological tissue/disorders, particularly applies to disorders involving these tissues and especially the tissues in which the gene is highly expressed. Moreover, since the gene is expressed in aortic intimal proliferations (atheroplaques), and heart tissue from patients with congestive heart failure, ischemia, and myopathy, the assays and methods involving pathology/disorders are particularly relevant in these disorders.

Since the 38911 gene is expressed in the tissues shown in Figures 12 and 13, the assays are particularly useful in cells derived from these tissue types, and particularly the cells and tissues in which the gene is highly expressed, such as spleen, tonsils, kidney, testis, liver, and osteoclasts. Furthermore, since the gene is expressed in these tissues, assays involving the protein in pathological tissue/disorders, particularly applies to disorders involving these tissues and especially the tissues in which the gene is highly expressed. Since the gene is highly expressed in osteoclasts, assays and methods involving pathology/disorders are particularly relevant to disorders involving osteoclast function. These disorders include but are not limited to those involved in bone growth and development, particularly disorders involving bone mass, such as osteoporosis. In addition, since relatively high expression occurs in fibrotic livers, liver fibrosis is a disorder relevant to expression of the 38911 receptor.

Further, expression of the 38911 receptor is relevant to inflammation, in view of homology to the C5a receptor.

Disorders involving the spleen include, but are not limited to, splenomegaly, including nonspecific acute splenitis, congestive spenomegaly, and spenic infarcts; neoplasms, congenital anomalies, and rupture. Disorders associated with splenomegaly include infections, such as nonspecific splenitis,

BNSDOCID: <WO__0149847A2_I_>

infectious mononucleosis, tuberculosis, typhoid fever, brucellosis, cytomegalovirus, syphilis, malaria, histoplasmosis, toxoplasmosis, kala-azar, trypanosomiasis, schistosomiasis, leishmaniasis, and echinococcosis; congestive states related to partial hypertension, such as cirrhosis of the liver, portal or splenic vein thrombosis, and cardiac failure; lymphohematogenous disorders, such as Hodgkin disease, non-Hodgkin lymphomas/leukemia, multiple myeloma, myeloproliferative disorders, hemolytic anemias, and thrombocytopenic purpura; immunologic-inflammatory conditions, such as rheumatoid arthritis and systemic lupus erythematosus; storage diseases such as Gaucher disease, Niemann-Pick disease, and mucopolysaccharidoses; and other conditions, such as amyloidosis, primary neoplasms and cysts, and secondary neoplasms.

Disorders involving the lung include, but are not limited to, congenital anomalies; atelectasis; diseases of vascular origin, such as pulmonary congestion and edema, including hemodynamic pulmonary edema and edema caused by microvascular injury, adult respiratory distress syndrome (diffuse alveolar damage), pulmonary embolism, hemorrhage, and infarction, and pulmonary hypertension and vascular sclerosis; chronic obstructive pulmonary disease, such as emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis; diffuse interstitial (infiltrative, restrictive) diseases, such as pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia (pulmonary infiltration with eosinophilia), *Bronchiolitis obliterans*-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, including Goodpasture syndrome, idiopathic pulmonary hemosiderosis and other hemorrhagic syndromes, pulmonary involvement in collagen vascular disorders, and pulmonary alveolar proteinosis; complications of therapies, such as drug-induced lung disease, radiation-induced lung disease, and lung transplantation; tumors, such as bronchogenic carcinoma, including paraneoplastic syndromes, bronchioloalveolar carcinoma, neuroendocrine tumors, such as bronchial carcinoid, miscellaneous tumors, and metastatic tumors; pathologies of the pleura, including inflammatory pleural effusions, noninflammatory pleural effusions, pneumothorax, and pleural tumors, including solitary fibrous tumors (pleural fibroma) and malignant mesothelioma.

Disorders involving the colon include, but are not limited to, congenital anomalies, such as atresia and stenosis, Meckel diverticulum, congenital aganglionic megacolon-Hirschsprung disease; enterocolitis, such as diarrhea and dysentery, infectious enterocolitis, including viral gastroenteritis, bacterial enterocolitis, necrotizing enterocolitis, antibiotic-associated colitis (pseudomembranous colitis), and collagenous and lymphocytic colitis, miscellaneous intestinal inflammatory disorders, including parasites and protozoa, acquired immunodeficiency syndrome, transplantation, drug-induced intestinal injury, radiation enterocolitis, neutropenic colitis (typhlitis), and diversion colitis; idiopathic inflammatory bowel disease, such as Crohn disease and ulcerative colitis; tumors of the colon, such as non-neoplastic polyps, adenomas, familial syndromes, colorectal carcinogenesis, colorectal carcinoma, and carcinoid tumors.

Disorders involving the liver include, but are not limited to, hepatic injury; jaundice and cholestasis, such as bilirubin and bile formation; hepatic failure and cirrhosis, such as cirrhosis, portal hypertension, including ascites, portosystemic shunts, and splenomegaly; infectious disorders, such as viral hepatitis, including hepatitis A-E infection and infection by other hepatitis viruses, clinicopathologic syndromes, such as the carrier state, asymptomatic infection, acute viral hepatitis, chronic viral hepatitis, and fulminant hepatitis; autoimmune hepatitis; drug- and toxin-induced liver disease, such as alcoholic liver disease; inborn errors of metabolism and pediatric liver disease, such as hemochromatosis, Wilson disease, a_1-antitrypsin deficiency, and neonatal hepatitis; intrahepatic biliary tract disease, such as secondary biliary cirrhosis, primary biliary cirrhosis, primary sclerosing cholangitis, and anomalies of the biliary tree; circulatory disorders, such as impaired blood flow into the liver, including hepatic artery compromise and portal vein obstruction and thrombosis, impaired blood flow through the liver, including passive congestion and centrilobular necrosis and peliosis hepatis, hepatic vein outflow obstruction, including hepatic vein thrombosis (Budd-Chiari syndrome) and veno-occlusive disease; hepatic disease associated with pregnancy, such as preeclampsia and eclampsia, acute fatty liver of pregnancy, and intrehepatic cholestasis of pregnancy; hepatic complications of organ or bone marrow transplantation, such as drug toxicity after bone marrow transplantation, graft-

BNSDOCID: <WO 0149847A2 | >

versus-host disease and liver rejection, and nonimmunologic damage to liver allografts; tumors and tumorous conditions, such as nodular hyperplasias, adenomas, and malignant tumors, including primary carcinoma of the liver and metastatic tumors.

Disorders involving the uterus and endometrium include, but are not limited to, endometrial histology in the menstrual cycle; functional endometrial disorders, such as anovulatory cycle, inadequate luteal phase, oral contraceptives and induced endometrial changes, and menopausal and postmenopausal changes; inflammations, such as chronic endometritis; adenomyosis; endometriosis; endometrial polyps; endometrial hyperplasia; malignant tumors, such as carcinoma of the endometrium; mixed Müllerian and mesenchymal tumors, such as malignant mixed Müllerian tumors; tumors of the myometrium, including leiomyomas, leiomyosarcomas, and endometrial stromal tumors.

Disorders involving the brain include, but are not limited to, disorders involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states--global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive cerebrovascular disease, including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain abscess, subdural empyema, and extradural abscess, chronic bacterial meningoencephalitis, including tuberculosis and mycobacterioses, neurosyphilis, and neuroborreliosis (Lyme disease), viral meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2,

Varicalla-zoster virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral neuropathy, and AIDS in children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute necrotizing hemorrhagic encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer disease and Pick disease, degenerative diseases of basal ganglia and brain stem, including Parkinsonism, idiopathic Parkinson disease (paralysis agitans), progressive supranuclear palsy, corticobasal degenration, multiple system atrophy, including striatonigral degenration, Shy-Drager syndrome, and olivopontocerebellar atrophy, and Huntington disease; spinocerebellar degenerations, including spinocerebellar ataxias, including Friedreich ataxia, and ataxia-telanglectasia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; inborn errors of metabolism, such as leukodystrophies, including Krabbe disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, and Canavan disease, mitochondrial encephalomyopathies, including Leigh disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic diseases, including vitamin deficiencies such as thiamine (vitamin B_1) deficiency and vitamin B_{12} deficiency, neurologic sequelae of metabolic disturbances, including hypoglycemia, hyperglycemia, and hepatic encephatopathy, toxic disorders, including carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate and radiation-induced injury; tumors, such as gliomas, including astrocytoma, including fibrillary (diffuse) astrocytoma and glioblastoma multiforme, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma, oligodendroglioma, and ependymoma and related paraventricular

EPSDOC'D: <WO 0149847A2 I >

mass lesions, neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors, including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including neurofibromotosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease.

Disorders involving T-cells include, but are not limited to, cell-mediated hypersensitivity, such as delayed type hypersensitivity and T-cell-mediated cytotoxicity, and transplant rejection; autoimmune diseases, such as systemic lupus erythematosus, Sjögren syndrome, systemic sclerosis, inflammatory myopathies, mixed connective tissue disease, and polyarteritis nodosa and other vasculitides; immunologic deficiency syndromes, including but not limited to, primary immunodeficiencies, such as thymic hypoplasia, severe combined immunodeficiency diseases, and AIDS; leukopenia; reactive (inflammatory) proliferations of white cells, including but not limited to, leukocytosis, acute nonspecific lymphadenitis, and chronic nonspecific lymphadenitis; neoplastic proliferations of white cells, including but not limited to lymphoid neoplasms, such as precursor T-cell neoplasms, such as acute lymphoblastic leukemia/lymphoma, peripheral T-cell and natural killer cell neoplasms that include peripheral T-cell lymphoma, unspecified, adult T-cell leukemia/lymphoma, mycosis fungoides and Sézary syndrome, and Hodgkin disease.

Diseases of the skin, include but are not limited to, disorders of pigmentation and melanocytes, including but not limited to, vitiligo, freckle, melasma, lentigo, nevocellular nevus, dysplastic nevi, and malignant melanoma; benign epithelial tumors, including but not limited to, seborrheic keratoses, acanthosis nigricans, fibroepithelial polyp, epithelial cyst, keratoacanthoma, and adnexal (appendage) tumors; premalignant and malignant epidermal tumors, including but not limited to, actinic keratosis, squamous cell carcinoma, basal cell

carcinoma, and merkel cell carcinoma; tumors of the dermis, including but not limited to, benign fibrous histiocytoma, dermatofibrosarcoma protuberans, xanthomas, and dermal vascular tumors; tumors of cellular immigrants to the skin, including but not limited to, histiocytosis X, mycosis fungoides (cutaneous T-cell lymphoma), and mastocytosis; disorders of epidermal maturation, including but not limited to, ichthyosis; acute inflammatory dermatoses, including but not limited to, urticaria, acute eczematous dermatitis, and erythema multiforme; chronic inflammatory dermatoses, including but not limited to, psoriasis, lichen planus, and lupus erythematosus; blistering (bullous) diseases, including but not limited to, pemphigus, bullous pemphigoid, dermatitis herpetiformis, and noninflammatory blistering diseases: epidermolysis bullosa and porphyria; disorders of epidermal appendages, including but not limited to, acne vulgaris; panniculitis, including but not limited to, erythema nodosum and erythema induratum; and infection and infestation, such as verrucae, molluscum contagiosum, impetigo, superficial fungal infections, and arthropod bites, stings, and infestations.

In normal bone marrow, the myelocytic series (polymorphoneuclear cells) make up approximately 60% of the cellular elements, and the erythrocytic series, 20-30%. Lymphocytes, monocytes, reticular cells, plasma cells and megakaryocytes together constitute 10-20%. Lymphocytes make up 5-15% of normal adult marrow. In the bone marrow, cell types are add mixed so that precursors of red blood cells (erythroblasts), macrophages (monoblasts), platelets (megakaryocytes), polymorphoneuclear leucocytes (myeloblasts), and lymphocytes (lymphoblasts) can be visible in one microscopic field. In addition, stem cells exist for the different cell lineages, as well as a precursor stem cell for the committed progenitor cells of the different lineages. The various types of cells and stages of each would be known to the person of ordinary skill in the art and are found, for example, on page 42 (Figure 2-8) of *Immunology, Imunopathology and Immunity*, Fifth Edition, Sell *et al.* Simon and Schuster (1996), incorporated by reference for its teaching of cell types found in the bone marrow. According, the invention is directed to disorders arising from these cells. These disorders include but are not limited to the following: diseases involving hematopoeitic stem

BNSDOCID: <WO 0149847A2 I >

cells; committed lymphoid progenitor cells; lymphoid cells including B and T-cells; committed myeloid progenitors, including monocytes, granulocytes, and megakaryocytes; and committed erythroid progenitors. These include but are not limited to the leukemias, including B-lymphoid leukemias, T-lymphoid leukemias, undifferentiated leukemias; erythroleukemia, megakaryoblastic leukemia, monocytic leukemia with and without differentiation; chronic and acute lymphoblastic leukemia, chronic and acute lymphocytic leukemia, chronic and acute myelogenous leukemia, lymphoma, myelo dysplastic syndrome, chronic and acute myeloid leukemia, myelomonocytic leukemia; chronic and acute myeloblastic leukemia, chronic and acute myelogenous leukemia, chronic and acute promyelocytic leukemia, chronic and acute myelocytic leukemia, hematologic malignancies of monocyte-macrophage lineage, such as juvenile chronic myelogenous leukemia; secondary AML, antecedent hematological disorder; refractory anemia; aplastic anemia; reactive cutaneous angioendotheliomatosis; fibrosing disorders involving altered expression in dendritic cells, disorders including systemic sclerosis, E-M syndrome, epidemic toxic oil syndrome, eosinophilic fasciitis localized forms of scleroderma, keloid, and fibrosing colonopathy; angiomatoid malignant fibrous histiocytoma; carcinoma, including primary head and neck squamous cell carcinoma; sarcoma, including kaposi's sarcoma; fibroadanoma and phyllodes tumors, including mammary fibroadenoma; stromal tumors; phyllodes tumors, including histiocytoma; erythroblastosis; neurofibromatosis; diseases of the vascular endothelium; demyelinating, particularly in old lesions; gliosis, vasogenic edema, vascular disease, Alzheimer's and Parkinson's disease; T-cell lymphomas; B-cell lymphomas.

Disorders involving the heart, include but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration

caused by calcification, such as calcific aortic stenosis, calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis and atresia, and aortic stenosis and atresia, and disorders involving cardiac transplantation.

Disorders involving blood vessels include, but are not limited to, responses of vascular cell walls to injury, such as endothelial dysfunction and endothelial activation and intimal thickening; vascular diseases including, but not limited to, congenital anomalies, such as arteriovenous fistula, atherosclerosis, and hypertensive vascular disease, such as hypertension; inflammatory disease--the vasculitides, such as giant cell (temporal) arteritis, Takayasu arteritis, polyarteritis nodosa (classic), Kawasaki syndrome (mucocutaneous lymph node syndrome), microscopic polyanglitis (microscopic polyarteritis, hypersensitivity or leukocytoclastic anglitis), Wegener granulomatosis, thromboanglitis obliterans (Buerger disease), vasculitis associated with other disorders, and infectious arteritis; Raynaud disease; aneurysms and dissection, such as abdominal aortic

ENSDOCID: <WO 0149847A2 I >

aneurysms, syphilitic (luetic) aneurysms, and aortic dissection (dissecting hematoma); disorders of veins and lymphatics, such as varicose veins, thrombophlebitis and phlebothrombosis, obstruction of superior vena cava (superior vena cava syndrome), obstruction of inferior vena cava (inferior vena cava syndrome), and lymphangitis and lymphedema; tumors, including benign tumors and tumor-like conditions, such as hemangioma, lymphangioma, glomus tumor (glomangioma), vascular ectasias, and bacillary angiomatosis, and intermediate-grade (borderline low-grade malignant) tumors, such as Kaposi sarcoma and hemangloendothelioma, and malignant tumors, such as angiosarcoma and hemangiopericytoma; and pathology of therapeutic interventions in vascular disease, such as balloon angioplasty and related techniques and vascular replacement, such as coronary artery bypass graft surgery.

Disorders involving red cells include, but are not limited to, anemias, such as hemolytic anemias, including hereditary spherocytosis, hemolytic disease due to erythrocyte enzyme defects: glucose-6-phosphate dehydrogenase deficiency, sickle cell disease, thalassemia syndromes, paroxysmal nocturnal hemoglobinuria, immunohemolytic anemia, and hemolytic anemia resulting from trauma to red cells; and anemias of diminished erythropoiesis, including megaloblastic anemias, such as anemias of vitamin B12 deficiency: pernicious anemia, and anemia of folate deficiency, iron deficiency anemia, anemia of chronic disease, aplastic anemia, pure red cell aplasia, and other forms of marrow failure.

Disorders involving the thymus include developmental disorders, such as DiGeorge syndrome with thymic hypoplasia or aplasia; thymic cysts; thymic hypoplasia, which involves the appearance of lymphoid follicles within the thymus, creating thymic follicular hyperplasia; and thymomas, including germ cell tumors, lynphomas, Hodgkin disease, and carcinoids. Thymomas can include benign or encapsulated thymoma, and malignant thymoma Type I (invasive thymoma) or Type II, designated thymic carcinoma.

Disorders involving B-cells include, but are not limited to precursor B-cell neoplasms, such as lymphoblastic leukemia/lymphoma. Peripheral B-cell neoplasms include, but are not limited to, chronic lymphocytic leukemia/small lymphocytic lymphoma, follicular lymphoma, diffuse large B-cell lymphoma,

Burkitt lymphoma, plasma cell neoplasms, multiple myeloma, and related entities, lymphoplasmacytic lymphoma (Waldenstrōm macroglobulinemia), mantle cell lymphoma, marginal zone lymphoma (MALToma), and hairy cell leukemia.

Disorders involving the kidney include, but are not limited to, congenital anomalies including, but not limited to, cystic diseases of the kidney, that include but are not limited to, cystic renal dysplasia, autosomal dominant (adult) polycystic kidney disease, autosomal recessive (childhood) polycystic kidney disease, and cystic diseases of renal medulla, which include, but are not limited to, medullary sponge kidney, and nephronophthisis-uremic medullary cystic disease complex, acquired (dialysis-associated) cystic disease, such as simple cysts; glomerular diseases including pathologies of glomerular injury that include, but are not limited to, in situ immune complex deposition, that includes, but is not limited to, anti-GBM nephritis, Heymann nephritis, and antibodies against planted antigens, circulating immune complex nephritis, antibodies to glomerular cells, cell-mediated immunity in glomerulonephritis, activation of alternative complement pathway, epithelial cell injury, and pathologies involving mediators of glomerular injury including cellular and soluble mediators, acute glomerulonephritis, such as acute proliferative (poststreptococcal, postinfectious) glomerulonephritis, including but not limited to, poststreptococcal glomerulonephritis and nonstreptococcal acute glomerulonephritis, rapidly progressive (crescentic) glomerulonephritis, nephrotic syndrome, membranous glomerulonephritis (membranous nephropathy), minimal change disease (lipoid nephrosis), focal segmental glomerulosclerosis, membranoproliferative glomerulonephritis, IgA nephropathy (Berger disease), focal proliferative and necrotizing glomerulonephritis (focal glomerulonephritis), hereditary nephritis, including but not limited to, Alport syndrome and thin membrane disease (benign familial hematuria), chronic glomerulonephritis, glomerular lesions associated with systemic disease, including but not limited to, systemic lupus erythematosus, Henoch-Schönlein purpura, bacterial endocarditis, diabetic glomerulosclerosis, amyloidosis, fibrillary and immunotactoid glomerulonephritis, and other systemic disorders; diseases affecting tubules and interstitium, including acute tubular necrosis and tubulointerstitial nephritis, including but not limited to, pyelonephritis

and urinary tract infection, acute pyelonephritis, chronic pyelonephritis and reflux nephropathy, and tubulointerstitial nephritis induced by drugs and toxins, including but not limited to, acute drug-induced interstitial nephritis, analgesic abuse nephropathy, nephropathy associated with nonsteroidal anti-inflammatory drugs, and other tubulointerstitial diseases including, but not limited to, urate nephropathy, hypercalcemia and nephrocalcinosis, and multiple myeloma; diseases of blood vessels including benign nephrosclerosis, malignant hypertension and accelerated nephrosclerosis, renal artery stenosis, and thrombotic microangiopathies including, but not limited to, classic (childhood) hemolytic-uremic syndrome, adult hemolytic-uremic syndrome/thrombotic thrombocytopenic purpura, idiopathic HUS/TTP, and other vascular disorders including, but not limited to, atherosclerotic ischemic renal disease, atheroembolic renal disease, sickle cell disease nephropathy, diffuse cortical necrosis, and renal infarcts; urinary tract obstruction (obstructive uropathy); urolithiasis (renal calculi, stones); and tumors of the kidney including, but not limited to, benign tumors, such as renal papillary adenoma, renal fibroma or hamartoma (renomedullary interstitial cell tumor), angiomyolipoma, and oncocytoma, and malignant tumors, including renal cell carcinoma (hypernephroma, adenocarcinoma of kidney), which includes urothelial carcinomas of renal pelvis.

Disorders of the breast include, but are not limited to, disorders of development; inflammations, including but not limited to, acute mastitis, periductal mastitis, periductal mastitis (recurrent subareolar abscess, squamous metaplasia of lactiferous ducts), mammary duct ectasia, fat necrosis, granulomatous mastitis, and pathologies associated with silicone breast implants; fibrocystic changes; proliferative breast disease including, but not limited to, epithelial hyperplasia, sclerosing adenosis, and small duct papillomas; tumors including, but not limited to, stromal tumors such as fibroadenoma, phyllodes tumor, and sarcomas, and epithelial tumors such as large duct papilloma; carcinoma of the breast including in situ (noninvasive) carcinoma that includes ductal carcinoma in situ (including Paget's disease) and lobular carcinoma in situ, and invasive (infiltrating) carcinoma including, but not limited to, invasive ductal carcinoma, no special type, invasive lobular carcinoma, medullary carcinoma,

colloid (mucinous) carcinoma, tubular carcinoma, and invasive papillary carcinoma, and miscellaneous malignant neoplasms.

Disorders in the male breast include, but are not limited to, gynecomastia and carcinoma.

Disorders involving the testis and epididymis include, but are not limited to, congenital anomalies such as cryptorchidism, regressive changes such as atrophy, inflammations such as nonspecific epididymitis and orchitis, granulomatous (autoimmune) orchitis, and specific inflammations including, but not limited to, gonorrhea, mumps, tuberculosis, and syphilis, vascular disturbances including torsion, testicular tumors including germ cell tumors that include, but are not limited to, seminoma, spermatocytic seminoma, embryonal carcinoma, yolk sac tumor choriocarcinoma, teratoma, and mixed tumors, tumore of sex cord-gonadal stroma including, but not limited to, leydig (interstitial) cell tumors and sertoli cell tumors (androblastoma), and testicular lymphoma, and miscellaneous lesions of tunica vaginalis.

Disorders involving the prostate include, but are not limited to, inflammations, benign enlargement, for example, nodular hyperplasia (benign prostatic hypertrophy or hyperplasia), and tumors such as carcinoma.

Disorders involving the thyroid include, but are not limited to, hyperthyroidism; hypothyroidism including, but not limited to, cretinism and myxedema; thyroiditis including, but not limited to, hashimoto thyroiditis, subacute (granulomatous) thyroiditis, and subacute lymphocytic (painless) thyroiditis; Graves disease; diffuse and multinodular goiter including, but not limited to, diffuse nontoxic (simple) goiter and multinodular goiter; neoplasms of the thyroid including, but not limited to, adenomas, other benign tumors, and carcinomas, which include, but are not limited to, papillary carcinoma, follicular carcinoma, medullary carcinoma, and anaplastic carcinoma; and cogenital anomalies.

Disorders involving the skeletal muscle include tumors such as rhabdomyosarcoma.

Disorders involving the pancreas include those of the exocrine pancreas such as congenital anomalies, including but not limited to, ectopic pancreas;

BNSDOCID: <WO___0149847A2_I_>

pancreatitis, including but not limited to, acute pancreatitis; cysts, including but not limited to, pseudocysts; tumors, including but not limited to, cystic tumors and carcinoma of the pancreas; and disorders of the endocrine pancreas such as, diabetes mellitus; islet cell tumors, including but not limited to, insulinomas, gastrinomas, and other rare islet cell tumors.

Disorders involving the small intestine include the malabsorption syndromes such as, celiac sprue, tropical sprue (postinfectious sprue), whipple disease, disaccharidase (lactase) deficiency, abetalipoproteinemia, and tumors of the small intestine including adenomas and adenocarcinoma.

Disorders related to reduced platelet number, thrombocytopenia, include idiopathic thrombocytopenic purpura, including acute idiopathic thrombocytopenic purpura, drug-induced thrombocytopenia, HIV-associated thrombocytopenia, and thrombotic microangiopathies: thrombotic thrombocytopenic purpura and hemolytic-uremic syndrome.

Disorders involving precursor T-cell neoplasms include precursor T lymphoblastic leukemia/lymphoma. Disorders involving peripheral T-cell and natural killer cell neoplasms include T-cell chronic lymphocytic leukemia, large granular lymphocytic leukemia, mycosis fungoides and Sézary syndrome, peripheral T-cell lymphoma, unspecified, angioimmunoblastic T-cell lymphoma, angiocentric lymphoma (NK/T-cell lymphoma[4a]), intestinal T-cell lymphoma, adult T-cell leukemia/lymphoma, and anaplastic large cell lymphoma.

Disorders involving the ovary include, for example, polycystic ovarian disease, Stein-leventhal syndrome, Pseudomyxoma peritonei and stromal hyperthecosis; ovarian tumors such as, tumors of coelomic epithelium, serous tumors, mucinous tumors, endometeriod tumors, clear cell adenocarcinoma, cystadenofibroma, brenner tumor, surface epithelial tumors; germ cell tumors such as mature (benign) teratomas, monodermal teratomas, immature malignant teratomas, dysgerminoma, endodermal sinus tumor, choriocarcinoma; sex cord-stomal tumors such as, granulosa-theca cell tumors, thecoma-fibromas, androblastomas, Hill cell tumors, and gonadoblastoma; and metastatic tumors such as Krukenberg tumors.

Bone-forming cells include the osteoprogenitor cells, osteoblasts, and osteocytes. The disorders of the bone are complex because they may have an impact on the skeleton during any of its stages of development. Hence, the disorders may have variable manifestations and may involve one, multiple or all bones of the body. Such disorders include, congenital malformations, achondroplasia and thanatophoric dwarfism, diseases associated with abnormal matix such as type 1 collagen disease, osteoporosis, Paget's disease, rickets, osteomalacia, high-turnover osteodystrophy, low-turnover of aplastic disease, osteonecrosis, pyogenic osteomyelitis, tuberculous osteomyelitis, osteoma, osteoid osteoma, osteoblastoma, osteosarcoma, osteochondroma, chondromas, chondroblastoma, chondromyxoid fibroma, chondrosarcoma, fibrous cortical defects, fibrous dysplasia, fibrosarcoma, malignant fibrous histiocytoma, Ewing's sarcoma, primitive neuroectodermal tumor, giant cell tumor, and metastatic tumors.

The polypeptides of the invention are useful for producing antibodies specific for the 39404, 38911, and 26904 protein, regions, or fragments. Regions having a high antigenicity index score are shown in Figures 3, 9, and 15.

The polypeptides, variants, and fragments (including those which may have been disclosed prior to the present invention) are useful for biological assays related to seven-transmembrane proteins/GPCRs. Such assays involve any of the known seven-transmembrane protein/GPCR functions or activities or properties useful for diagnosis and treatment of seven-transmembrane protein/GPCR-related conditions.

The polypeptides of the invention are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the protein, as a biopsy or expanded in cell culture. In one embodiment, however, cell-based assays involve recombinant host cells expressing the protein.

Determining the ability of the test compound to interact with the polypeptide can also comprise determining the ability of the test compound to preferentially bind to the polypeptide as compared to the ability of the ligand, or a biologically active portion thereof, to bind to the polypeptide.

BNSDOCID: <WO___0149847A2_I_>

The polypeptides can be used to identify compounds that modulate protein activity. Such compounds, for example, can increase or decrease affinity or rate of binding to a known ligand, compete with ligand for binding to the protein, or displace ligand bound to the protein. 39404, 38911, and 26904 protein and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the protein. These compounds can be further screened against a functional protein to determine the effect of the compound on the protein activity. Compounds can be identified that activate (agonist) or inactivate (antagonist) the protein to a desired degree. Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). Examples for the 39404 protein include but are not limited to purine analogs such as those discussed above. Examples for the 38911 protein include but are not limited to C5a and C5a analogs.

The polypeptides of the invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the protein and a target molecule that normally interacts with the protein. The target can be ligand or a component of the signal pathway with which the protein normally interacts (for example, a G-protein or other interactor involved in cAMP or phosphatidylinositol turnover and/or adenylate cyclase, or phospholipase C activation). The assay includes the steps of combining the protein with a candidate compound under conditions that allow the protein or fragment to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the protein and the target, such as any of the associated effects of signal transduction such as G-protein phosphorylation, cyclic AMP or phosphatidylinositol turnover, and adenylate cyclase or phospholipase C activation.

Determining the ability of the protein to bind to a target molecule can also be accomplished using a technology such as real-time Bimolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem. 63*:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol. 5*:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without

labeling any of the interactants (e.g., BIAcore™). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des. 12*:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. USA 90*:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA 91*:11422; Zuckermann *et al.* (1994). *J. Med. Chem. 37*:2678; Cho *et al.* (1993) *Science 261*:1303; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl. 33*:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl. 33*:2061; and in Gallop *et al.* (1994) *J. Med. Chem. 37*:1233. Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques 13*:412-421), or on beads (Lam (1991) *Nature 354*:82-84), chips (Fodor (1993) *Nature 364*:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. USA 89*:1865-1869) or on phage (Scott and Smith (1990) *Science 249*:386-390); (Devlin (1990) *Science 249*:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci. 97*:6378-6382); (Felici (1991) *J. Mol. Biol. 222*:301-310); (Ladner *supra*).

Candidate compounds include, for example, 1) purine analogs (39404), 2) peptides such as soluble peptides, including C5a (38911), C5a fragments, and derivatives thereof, Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al., Nature 354*:82-84 (1991); Houghten *et al., Nature 354*:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 3) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g.,

BNSDOCID: <WO__0149847A2_I_>

Songyang *et al., Cell* 72:767-778 (1993)); 4) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, $F(ab')_2$, Fab expression library fragments, and epitope-binding fragments of antibodies); and 5) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble full-length protein or fragment that competes for ligand binding. Other candidate compounds include mutant proteins or appropriate fragments containing mutations that affect protein function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention provides other end points to identify compounds that modulate (stimulate or inhibit) receptor activity. The assays typically involve an assay of events in the signal transduction pathway that indicate receptor activity. Thus, the expression of genes that are up- or down-regulated in response to the receptor protein dependent signal cascade can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase. Alternatively, phosphorylation of the protein, or a protein target, could also be measured.

Binding and/or activating compounds can also be screened by using chimeric proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a G-protein-binding region can be used that interacts with a different G-protein then that which is recognized by the native receptor. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. Alternatively, the entire transmembrane portion or subregions (such as transmembrane segments or intracellular or extracellular loops) can be replaced with the entire transmembrane portion or subregions specific to a host cell that is different from the host cell from which the amino terminal extracellular domain and/or the G-protein-binding

region are derived. This allows for assays to be performed in other than the specific host cell from which the protein is derived. Alternatively, the amino terminal extracellular domain (and/or other ligand-binding regions) could be replaced by a domain (and/or other binding region) binding a different ligand, thus, providing an assay for test compounds that interact with the heterologous amino terminal extracellular domain (or region) but still cause signal transduction. Finally, activation can be detected by a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native signal transduction pathway.

The polypeptides of the invention are also useful in competition binding assays in methods designed to discover compounds that interact with the protein. Thus, a compound is exposed to a polypeptide of the invention under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble polypeptide is also added to the mixture. If the test compound interacts with the soluble polypeptide, it decreases the amount of complex formed or activity from the protein target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the protein. Thus, the soluble polypeptide that competes with the target region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is desirable to immobilize either the protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/39404, 38911, and 26904 fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to

BNSDOCID: <WO___0149847A2_I_>

remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of receptor-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a protein of the invention-binding protein and a candidate compound are incubated in the protein of the invention-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the protein target molecule, or which are reactive with protein and compete with the target molecule; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Modulators of 39404 protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the protein pathway, by treating cells that express the 39404 protein, such as in breast, brain, kidney, vein, fetal kidney, fetal liver, aortic intimal proliferations, internal mammary artery, and cells involved in congestive heart failure, ischemia, and myopathy, for example, cardiomyocytes. Modulators of 38911 protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the protein pathway, by treating cells that express the 38911 protein, such as in Figures 12 and 13, and especially osteoclasts, liver, kidney, and testis. Modulators of 39604 protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the protein pathway, by treating cells that express the 26904 protein, such as in brain.

Treatment is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease.

A therapeutic agent or compound includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

The polypeptides of the invention are thus useful for treating a protein of the invention-associated disorder characterized by aberrant expression or activity of a protein of the invention . In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) expression or activity of the protein. In another embodiment, the method involves administering a protein as therapy to compensate for reduced or aberrant expression or activity of the protein.

Stimulation of protein activity is desirable in situations in which the protein is abnormally downregulated and/or in which increased protein activity is likely to have a beneficial effect. Likewise, inhibition of protein activity is desirable in situations in which the protein is abnormally upregulated and/or in which decreased protein activity is likely to have a beneficial effect. In one example of such a situation, a subject has a disorder characterized by aberrant development or cellular differentiation. In another example of such a situation, the subject has a proliferative disease (e.g., cancer) or a disorder characterized by an aberrant hematopoietic response. In another example of such a situation, it is desirable to achieve tissue regeneration in a subject (e.g., where a subject has undergone brain or spinal cord injury and it is desirable to regenerate neuronal tissue in a regulated manner).

In yet another aspect of the invention, the proteins of the invention can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell 72*:223-232; Madura *et al.* (1993) *J. Biol. Chem. 268*:12046-12054; Bartel *et al.* (1993) *Biotechniques 14*:920-924;

BNSDOCID: <WO___0149847A2_I_>

Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO 94/10300), to identify other proteins (captured proteins) which bind to or interact with the proteins of the invention and modulate their activity.

The 39404 polypeptides also are useful to provide a target for diagnosing a disease or predisposition to disease mediated by the protein, especially in breast, brain, kidney, vein, fetal kidney, fetal liver, aortic intimal proliferations, internal mammary artery, and especially in congestive heart failure, ischemia, and myopathy. Disorders, however, also include diseases of other tissues in which the gene is expressed as shown in Figures 5-7. Tissue disorders are described in more detail hereinabove. The 38911 polypeptides also are useful to provide a target for diagnosing a disease or predisposition to disease mediated by the protein, especially in osteoclasts, liver, kidney, and testis. The 26904 polypeptides also are useful to provide a target for diagnosing a disease or predisposition to disease mediated by the protein, such as in brain. Accordingly, methods are provided for detecting the presence, or levels of, the protein in a cell, tissue, or organism. The method involves contacting a biological sample with a compound capable of interacting with the protein such that the interaction can be detected.

One agent for detecting a protein of the invention is an antibody capable of selectively binding to the protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The proteins of the invention also provide a target for diagnosing active disease, or predisposition to disease, in a patient having a variant protein. Thus, a protein of the invention can be isolated from a biological sample, assayed for the presence of a genetic mutation that results in an aberrant protein. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered protein activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein.

In vitro techniques for detection of the protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. Alternatively, the protein can be detected *in vivo* in a subject by introducing into the subject a labeled antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods which detect the allelic variant of a protein of the invention expressed in a subject and methods which detect fragments of a protein of the invention in a sample.

The polypeptides of the invention are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M., *Clin. Exp. Pharmacol. Physiol. 23(10-11)*:983-985 (1996), and Linder, M.W., *Clin. Chem. 43(2)*:254-266 (1997). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the protein in which one or more of the protein functions in one population is different from those in another population. The polypeptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to

SNSDOCID: <WO 0149847A2 | >

amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and receptor activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic polypeptides could be identified.

The polypeptides of the invention are also useful for monitoring therapeutic effects during clinical trials and other treatment. Thus, the therapeutic effectiveness of an agent that is designed to increase or decrease gene expression, protein levels or activity can be monitored over the course of treatment using the polypeptides as an end-point target. The monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression or activity of a specified protein in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the protein in the post-administration samples; (v) comparing the level of expression or activity of the protein in the pre-administration sample with the protein in the post-administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

The polypeptides of the invention are also useful for treating an associated disorder. Accordingly, methods for treatment include the use of soluble protein or fragments of the protein that compete for ligand binding. These proteins or fragments can have a higher affinity for the ligand so as to provide effective competition.

Antibodies

The invention also provides antibodies that selectively bind to the 39404, 38911, and 26904 proteins and variants and fragments. An antibody is considered to selectively bind, even if it also binds to other proteins that are not substantially homologous with the proteins. These other proteins share homology with a fragment or domain of the protein of the invention . This conservation in specific regions gives rise to antibodies that bind to both proteins by virtue of the

homologous sequence. In this case, it would be understood that antibody binding to the protein of the invention is still selective.

To generate antibodies, an isolated polypeptide of the invention is used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. Either the full-length protein or antigenic peptide fragment can be used. Regions having a high antigenicity index are shown in Figures 3, 9, and 15.

Antibodies are preferably prepared from these regions or from discrete fragments in these regions. However, antibodies can be prepared from any region of the peptide as described herein. A preferred fragment produces an antibody that diminishes or completely prevents ligand-binding. Antibodies can be developed against the entire protein or portions of the protein, for example, the intracellular carboxy terminal domain, the amino terminal extracellular domain, the entire transmembrane domain or specific segments, any of the intra or extracellular loops, or any portions of the above. Antibodies may also be developed against specific functional sites, such as the site of ligand-binding, the site of G protein coupling, or sites that are phosphorylated, glycosylated, or myristoylated.

An antigenic 39404, 38911, or 26904 fragment will typically comprise at least 8-10 contiguous amino acid residues. The antigenic peptide can comprise, however, a contiguous sequence of at least 12, 14 amino acid residues, at least 15 amino acid residues, at least 20 amino acid residues, or at least 30 amino acid residues. In one embodiment, fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions. These fragments are not to be construed, however, as encompassing any fragments which may be disclosed prior to the invention.

Antibodies can be polyclonal or monoclonal. An intact antibody, or a fragment thereof (e.g. Fab or $F(ab')_2$) can be used.

Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or

BNSDOCID: <WO___0149847A2_I_>

acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

An appropriate immunogenic preparation can be derived from native, recombinantly expressed, protein or chemically synthesized peptides.

Antibody Uses

The antibodies can be used to isolate a protein of the invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells.

The antibodies are useful to detect the presence of a protein of the invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development.

The antibodies can be used to detect a protein of the invention *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression.

The antibodies can be used to assess abnormal tissue distribution or abnormal expression during development.

Antibody detection of circulating fragments of the full length protein can be used to identify protein turnover.

Further, the antibodies can be used to assess expression of a protein of the invention in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to protein function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, or level of expression of the protein, the antibody can be prepared against the normal protein. If a disorder is characterized by a specific mutation in the protein,

BNSDOCID: <WO___0149847A2_I_>

antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Antibodies can be developed against the whole protein or portions of the receptor, for example, portions of the amino terminal extracellular domain or extracellular loops.

The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting protein expression level or the presence of aberrant proteins of the invention and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Antibodies accordingly can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins of the invention can be used to identify individuals that require modified treatment modalities.

The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful in forensic identification. Accordingly, where an individual has been correlated with a specific genetic polymorphism resulting in a specific polymorphic protein, an antibody specific for the polymorphic protein can be used as an aid in identification.

The antibodies are also useful for inhibiting protein function, for example, blocking ligand binding.

These uses can also be applied in a therapeutic context in which treatment involves inhibiting protein function. An antibody can be used, for example, to

BNSDOCID: <WO___0149847A2_I_>

block ligand binding. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein associated with a cell.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol. 13*:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, e.g., U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806.

The invention also encompasses kits for using antibodies to detect the presence of a protein of the invention in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting the protein in a biological sample; means for determining the amount of the protein in the sample; and means for comparing the amount of the protein in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect the protein.

Polynucleotides

The nucleotide sequence in SEQ ID NO:2 was obtained by sequencing the deposited human full length cDNA. Accordingly, the sequence of the deposited clone is controlling as to any discrepancies between the two and any reference to the sequence of SEQ ID NO:2 includes reference to the sequence of the deposited cDNA.

The nucleotide sequence in SEQ ID NO:4 was obtained by sequencing the deposited human full length cDNA. Accordingly, the sequence of the deposited clone is controlling as to any discrepancies between the two and any reference to the sequence of SEQ ID NO:4 includes reference to the sequence of the deposited cDNA.

The nucleotide sequence in SEQ ID NO:6 was obtained by sequencing the deposited human full length cDNA. Accordingly, the sequence of the deposited

BNSDOCID: <WO___0149847A2_I_>

clone is controlling as to any discrepancies between the two and any reference to the sequence of SEQ ID NO:6 includes reference to the sequence of the deposited cDNA.

The specifically disclosed cDNAs comprise the coding region and 5′ and 3′ untranslated sequences (SEQ ID NO:2, SEQ ID NO:4, and SEQ ID NO:6).

The human 39404 cDNA is approximately 1729 nucleotides in length and encodes a full length protein that is approximately 337 amino acid residues in length. The nucleic acid is expressed in the tissues shown in Figures 5-7. Structural analysis of the amino acid sequence of SEQ ID NO:1 is provided in Figure 2, a hydropathy plot. The figure shows the putative structure of the seven transmembrane segments, the amino terminal extracellular domain and the carboxy terminal intracellular domain.

The human 38911 cDNA is approximately 1334 nucleotides in length and encodes a full length protein that is approximately 337 amino acid residues in length. The nucleic acid is expressed in the tissues shown in Figures 12 and 13. Structural analysis of the amino acid sequence of SEQ ID NO:3 is provided in Figure 9, a hydropathy plot. The figure shows the putative structure of the seven transmembrane segments, the amino terminal extracellular domain and the carboxy terminal intracellular domain.

The human 26904 cDNA is approximately 1743 nucleotides in length and encodes a full length protein that is approximately 450 amino acid residues in length. Structural analysis of the amino acid sequence of SEQ ID NO:5 is provided in Figure 14, a hydropathy plot. The figure shows the putative structure of the seven transmembrane segments, the amino terminal extracellular domain and the carboxy terminal intracellular domain.

As used herein, the term "transmembrane segment" refers to a structural amino acid motif which includes a hydrophobic helix that spans the plasma membrane. The entire transmembrane domain of 39404 spans from about amino acid 38 to about amino acid 305. The entire transmembrane domain of 38911 spans from about amino acid 41 to about amino acid 294. The entire transmembrane domain of 26904 spans from about amino acid 30 to about amino acid 430.

BNSDOCID: <WO___0149847A2_I_>

The invention provides isolated polynucleotides encoding a 39404 protein. The term "39404 polynucleotide" or "39404 nucleic acid" refers to the sequence shown in SEQ ID NO:2 or in the deposited cDNA.

The invention provides isolated polynucleotides encoding a 38911 protein. The term "38911 polynucleotide" or "38911 nucleic acid" refers to the sequence shown in SEQ ID NO:4 or in the deposited cDNA.

The invention provides isolated polynucleotides encoding a 26904 protein. The term "26904 polynucleotide" or "26904 nucleic acid" refers to the sequence shown in SEQ ID NO:6 or in the deposited cDNA.

The term "polynucleotide" or "nucleic acid" further includes variants and fragments of the 39404, 38911, and 26904 polynucleotides.

An "isolated" nucleic acid of the invention is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5′ and 3′ ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB. The important point is that the nucleic acid is isolated from flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences of the invention.

Moreover, an "isolated" nucleic acid molecule, such as a cDNA or RNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the

present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 % (on a molar basis) of all macromolecular species present.

The polynucleotides of the invention can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

The polynucleotides of the invention include, but are not limited to, the sequence encoding the mature polypeptide alone, the sequence encoding the mature polypeptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature polypeptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5′ and 3′ sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the polynucleotide may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Polynucleotides of the invention can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded.

Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

One nucleic acid comprises the nucleotide sequence shown in SEQ ID NO:2, corresponding to human 39404 cDNA.

One nucleic acid comprises the nucleotide sequence shown in SEQ ID NO:4, corresponding to human 38911 cDNA.

One nucleic acid comprises the nucleotide sequence shown in SEQ ID NO:6, corresponding to human 26904 cDNA.

In one embodiment, the nucleic acid comprises only the coding region.

The invention further provides variant polynucleotides, and fragments thereof, that differ from the nucleotide sequence shown in SEQ ID NO:2 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence shown in SEQ ID NO:2.

The invention further provides variant polynucleotides, and fragments thereof, that differ from the nucleotide sequence shown in SEQ ID NO:4 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence shown in SEQ ID NO:4.

The invention further provides variant polynucleotides, and fragments thereof, that differ from the nucleotide sequence shown in SEQ ID NO:6 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence shown in SEQ ID NO:6.

The invention also provides nucleic acid molecules encoding the variant polypeptides described herein. Such polynucleotides may be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions.

Typically, variants have a substantial identity with a nucleic acid molecule selected from the group consisting of SEQ ID NOS:2, 4, 6, and 8 and the complements thereof.

Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. 39404 variants comprise a nucleotide sequence encoding a protein that is 40-45%, 45-50%, 50-55%, 55-60%, typically at least about 60-65%, 65-70%, or 70-75%, more typically at least about 70-75%, 75-80%, or 80-85%, and most typically at least about 85-90% or 90-95% or more homologous to the nucleotide sequence shown in SEQ ID NO:2 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO:2 or a fragment of the sequence.

38911 variants comprise a nucleotide sequence encoding a protein that is 35-40%, 40-45%, 45-50%, 50-55%, 55-60%, 60-65%, 65-70%, typically at least about 70-75%, more typically at least about 75-80% or 80-85%, and most typically at least about 85-90% or 90-95% or more homologous to the nucleotide sequence shown in SEQ ID NO:4 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO:4 or a fragment of the sequence.

26904 variants comprise a nucleotide sequence encoding a protein that is 50-55%, 55-60%, 60-65%, 65-70%, typically at least about 70-75%, more typically at least about 75-80% or 80-85%, and most typically at least about 85-90% or 90-95% or more homologous to the nucleotide sequence shown in SEQ ID NO:6 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO:6 or a fragment of the sequence.

It is understood that stringent hybridization does not indicate substantial homology where it is due to general homology, such as poly A sequences, or sequences common to all or most proteins, all seven-transmembrane proteins, all GPCRs, or all family I GPCRs or even all purinoceptors or C5a receptors.

BNSDOCID: <WO___0149847A2_I_>

Moreover, it is understood that variants do not include any of the nucleic acid sequences that may have been disclosed prior to the invention.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a polypeptide at least 50-55%, 55% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 90%, at least about 95% or more identical to each other remain hybridized to one another. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6, incorporated by reference. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45EC, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65EC. In another non-limiting example, nucleic acid molecules are allowed to hybridize in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more low stringency washes in 0.2X SSC/0.1% SDS at room temperature, or by one or more moderate stringency washes in 0.2X SSC/0.1% SDS at 42°C, or washed in 0.2X SSC/0.1% SDS at 65°C for high stringency. In one embodiment, an isolated nucleic acid molecule that hybridizes under stringent conditions to the sequence of SEQ ID NOS:2, 4, 6, or 8 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein). Timing of hybridization can vary from ½ hour to 10 hours or longer. Shorter hybridizations however can include from 1 to 5, and 6 to 10 hours. Typically, hybridization is performed overnight for around 10-12 hours. The time of washes can also vary from around 10 minutes to 30 minutes. Typically, washes are performed from 10-20 minutes.

As understood by those of ordinary skill, the exact conditions can be determined empirically and depend on ionic strength, temperature and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS. Other factors considered in determining the desired hybridization

conditions include the length of the nucleic acid sequences, base composition, percent mismatch between the hybridizing sequences and the frequency of occurrence of subsets of the sequences within other non-identical sequences. Thus, equivalent conditions can be determined by varying one or more of these parameters while maintaining a similar degree of identity or similarity between the two nucleic acid molecules.

The present invention also provides isolated nucleic acids that contain a single or double stranded fragment or portion that hybridizes under stringent conditions to a nucleotide sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, or 8 and the complements of SEQ ID NOS:2, 4, 6, or 8. In one embodiment, the nucleic acid consists of a portion of a nucleotide sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, or 8 and the complements SEQ ID NOS:2, 4, 6, or 8. The nucleic acid fragments of the invention are at least about 15, preferably at least about 18, 20, 23 or 25 nucleotides, and can be 30, 40, 50, 100, 200 or more nucleotides in length. Longer fragments, for example, 30 or more nucleotides in length, which encode antigenic proteins or polypeptides described herein are useful.

Furthermore, the invention provides polynucleotides that comprise a fragment of the full length polynucleotides of the invention. The fragment can be single or double stranded and can comprise DNA or RNA. The fragment can be derived from either the coding or the non-coding sequence.

In one embodiment, an isolated 39404 nucleic acid is at least 23 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2. The isolated fragments can be at least between 5-10, 10-20, 20-30, 30-40, 40-50, etc. including but not limited to 50, 75, 100, 200, 250, or 500 nucleotides in length or greater.

In another embodiment, an isolated 38911 nucleic acid from around nucleotide 1 to around nucleotide 200 is at least 5 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:4. In other embodiments, the isolated nucleic acid is from around nucleotide 950 to nucleotide 1080 and is at least five nucleotides in length, hybridizing under stringent conditions. In other

embodiments, from about nucleotide 190 to about nucleotide 950, fragments can be at least 5-10 nucleotides, at least 10-15 nucleotides, at least 15-20 nucleotides, at least 20-25 nucleotides, at least 25-30 nucleotides, at least 30-35 nucleotides, at least 35-40 nucleotides, for example, greater than 13 nucleotides, greater than 14 nucleotides, and greater than 18 nucleotides. In other embodiments, the nucleic acid is at least 40, 50, 100, 250, or 500 nucleotides in length or greater.

In another embodiment, an isolated 26904 nucleic acid from nucleotide 1 to around nucleotide 498 is at least 14 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:6. In another embodiment, the nucleic acid from around nucleotide 691 to around 1014 is at least 14 nucleotides. In other embodiments, the nucleic acid is at least 40, 50, 100, 250, or 500 nucleotides in length or greater.

In another embodiment, an isolated 39404 nucleic acid encodes the entire coding region from amino acid 1 to amino acid 337. In another embodiment, the isolated 39404 nucleic acid encodes a sequence corresponding to the mature protein from about amino acid 6 to about amino acid 337. In another embodiment, an isolated 38911 nucleic acid encodes the entire coding region from amino acid 1 to amino acid 337. In another embodiment, the isolated 38911 nucleic acid encodes a sequence corresponding to the mature protein from about amino acid 6 to about amino acid 337. In another embodiment, an isolated 26904 nucleic acid encodes the entire coding region from amino acid 1 to amino acid 450. In another embodiment, the isolated 26904 nucleic acid encodes a sequence corresponding to the mature protein from about amino acid 6 to about amino acid 450.

Other fragments of all four proteins include nucleotide sequences encoding the amino acid fragments described herein. Further, fragments can include subfragments of the specific domains or sites described herein. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments, according to the present invention, are not to be construed as encompassing those fragments that may have been disclosed prior to the invention except as they are used in methods involving tissues/disorders with which gene expression is associated.

However, it is understood that a nucleic acid fragment includes any nucleic acid sequence that does not include the entire gene.

39404 nucleic acid fragments further include sequences corresponding to the domains described herein, subregions also described, and specific functional sites. 39404 nucleic acid fragments include but are not limited to nucleic acid molecules encoding a polypeptide comprising the amino terminal extracellular domain, comprising the region spanning the transmembrane domain, a polypeptide comprising the carboxy terminal intracellular domain, and a polypeptide encoding the G-protein receptor signature (130-132 or surrounding amino acid residues from about 120 to about 140), nucleic acid molecules encoding any of the seven transmembrane segments, extracellular or intracellular loops, glycosylation sites or phosphorylation sites.

38911 nucleic acid fragments include but are not limited to nucleic acid molecules encoding a polypeptide comprising the amino terminal extracellular domain, the region spanning the transmembrane domain, and/or the carboxy terminal intracellular domain, and nucleic acid molecules encoding any of the seven transmembrane segments, extracellular or intracellular loops, glycosylation sites and phosphorylation sites.

26904 nucleic acid fragments include but are not limited to nucleic acid molecules encoding a polypeptide comprising the amino terminal extracellular domain, a polypeptide comprising the region spanning the transmembrane domain, and/or the carboxy terminal intracellular domain, and nucleic acid molecules encoding any of the seven transmembrane segments, extracellular or intracellular loops, glycosylation sites, protein kinase C, cAMP, cGMP, and casein kinase II phosphorylation sites, and myristoylation sites.

Where the location of the domains have been predicted by computer analysis, one of ordinary skill would appreciate that the amino acid residues constituting these domains can vary depending on the criteria used to define the domains.

Nucleic acid fragments also include combinations of the domains, segments, loops, and other functional sites described above. Thus, for example, a nucleic acid could include sequences corresponding to the amino terminal

BNSDOCID <WO___0149847A2_I_>

extracellular domain and one transmembrane segment. A person of ordinary skill in the art would be aware of the many permutations that are possible.

Where the location of the domains or sites have been predicted by computer analysis, one of ordinary skill would appreciate that the amino acid residues constituting these domains can vary depending on the criteria used to define the domains.

The invention also provides nucleic acid fragments that encode epitope bearing regions of the proteins described herein.

The isolated polynucleotide sequences, and especially fragments, are useful as DNA probes and primers.

For example, the coding region of a gene of the invention can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of the genes of the invention.

A probe/primer typically comprises substantially purified oligonucleotide. The 39404 oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 10, 20, typically about 25, more typically about 40, 50 or 75 consecutive nucleotides of SEQ ID NO:2 sense or anti-sense strand or other receptor polynucleotides. The 38911 oligonucleotide typically comprises a region of nucleotide sequence from 1-1080 that hybridizes under stringent conditions to at least about 15, typically about 25, more typically about 40, 50 or 75 consecutive nucleotides of SEQ ID NO:4 sense or anti-sense strand or other polynucleotides. The 26904 oligonucleotide typically comprises a region of nucleotide sequence from 1-498 of at least about 14, typically about 25, more typically about 40, 50 or 75 consecutive nucleotides of SEQ ID NO:6 sense or anti-sense strand or other polynucleotides that hybridizes under stringent conditions. The 26904 oligonucleotide also typically comprises a region of nucleotide sequence from nucleotide 691-1014 at least about 14, typically about 25, more typically about 40, 50, or 75 consecutive nucleotides of

SEQ ID NO:6 sense or anti-sense strand or other polynucleotides that hybridizes under stringent conditions.

Polynucleotide Uses

The nucleic acid sequences of the present invention can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol. 215*:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res. 25*(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The nucleic acid fragments of the invention provide probes or primers in assays such as those described below. "Probes" are oligonucleotides that hybridize in a base-specific manner to a complementary strand of nucleic acid. Such probes include polypeptide nucleic acids, as described in Nielsen *et al.* (1991) *Science 254*:1497-1500. Typically, a probe comprises a region of nucleotide sequence that hybridizes under highly stringent conditions to at least about 15, typically about 20-25, and more typically about 40, 50 or 75 consecutive nucleotides of a nucleic acid selected from the group consisting of SEQ ID NOS:2, 4, 6 or 8 and the complements thereof. More typically, the probe further comprises a label, e.g., radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

As used herein, the term "primer" refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis using well-known methods (e.g., PCR, LCR) including, but not limited to those described herein. The appropriate length of the primer depends on the particular use, but typically ranges from about 15 to 30 nucleotides. The term "primer site" refers to the area of the target DNA to which a primer hybridizes.

BNSDOCID: <WO___0149847A2_I_>

The term "primer pair" refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5' end of the nucleic acid sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the sequence to be amplified.

The polynucleotides are useful for various biological assays as described in detail below. Since the 39404 gene is expressed in the tissues shown in Figures 5-7, the assays are particularly useful in cells derived from these tissue types, and particularly the tissues in which the gene is highly expressed, such as brain, kidney, fetal kidney, fetal liver, internal mammary artery, and aortic intimal proliferations. Furthermore, since the gene is expressed in these tissues, assays involving the polynucleotides in pathological tissue/disorders, particularly applies to disorders involving these tissues and especially the tissues in which the gene is highly expressed. Moreover, since the gene is expressed in aortic intimal proliferations (atheroplaques), and heart tissue from patients with congestive heart failure, ischemia, and myopathy, the assays and methods involving pathology/disorders are particularly relevant in these disorders.

Since the 38911 is expressed in the tissues shown in Figures 12 and 13, the assays are particularly useful in cells derived from these tissue types, and particularly the tissues in which the gene is highly expressed, such as kidney, spleen, fibrotic liver tissue, tonsils, osteoclasts, liver, and testis. Furthermore, since the gene is expressed in these tissues, assays involving the polynucleotides and pathological tissues/disorders, particularly applies to disorders involving these tissues and especially the tissues in which the gene is highly expressed. Moreover, since the gene is expressed in liver fibrosis, the assays and methods involving pathology/disorders are particularly relevant in these disorders. Finally, in view of the fact that the gene is highly expressed in osteoclasts, assays and methods involving osteoporosis are particularly relevant.

The receptor polynucleotides are useful for probes, primers, and in biological assays.

Where the polynucleotides are used to assess seven-transmembrane protein/GPCR properties or functions, such as in the assays described herein, all or less than all of the entire cDNA can be useful. In this case, even fragments that

BNSDOCID: <WO___0149847A2_I_>

may have been known prior to the invention are encompassed. Thus, for example, assays specifically directed to seven-transmembrane protein/GPCR functions, such as assessing agonist or antagonist activity, encompass the use of known fragments. Further, diagnostic methods for assessing protein function can also be practiced with any fragment, including those fragments that may have been known prior to the invention. Similarly, in methods involving treatment of protein dysfunction, all fragments are encompassed including those which may have been known in the art.

The 39404 polynucleotides are useful as a hybridization probe for cDNA and genomic DNA to isolate a full-length cDNA and genomic clones encoding the polypeptide described in SEQ ID NO:1 and to isolate cDNA and genomic clones that correspond to variants producing the same polypeptide shown in SEQ ID NO:1 or the other variants described herein. Variants can be isolated from the same tissue and organism from which the polypeptide shown in SEQ ID NO:1 was isolated, different tissues from the same organism, or from different organisms.

The 38911 polynucleotides are useful as a hybridization probe for cDNA and genomic DNA to isolate a full-length cDNA and genomic clones encoding the polypeptide described in SEQ ID NO:3 and to isolate cDNA and genomic clones that correspond to variants producing the same polypeptide shown in SEQ ID NO:3 or the other variants described herein. Variants can be isolated from the same tissue and organism from which the polypeptide shown in SEQ ID NO:3 was isolated, different tissues from the same organism, or from different organisms.

The 26904 polynucleotides are useful as a hybridization probe for cDNA and genomic DNA to isolate a full-length cDNA and genomic clones encoding the polypeptide described in SEQ ID NO:5 and to isolate cDNA and genomic clones that correspond to variants producing the same polypeptide shown in SEQ ID NO:5 or the other variants described herein. Variants can be isolated from the same tissue and organism from which the polypeptide shown in SEQ ID NO:5 was isolated, different tissues from the same organism, or from different organisms.

BNSDOCID: <WO___0149847A2_I_>

This method is useful for isolating genes and cDNA that are developmentally-controlled and therefore may be expressed in the same tissue or different tissues at different points in the development of an organism.

The probe can correspond to any sequence along the entire length of the gene encoding the protein. Accordingly, it could be derived from 5′ noncoding regions, the coding region, and 3′ noncoding regions. Probes, however, are not to be construed as corresponding to any sequences that may be known prior to the invention.

The 39404 nucleic acid probe can be, for example, the full-length cDNA of SEQ ID NO:1, or a fragment thereof, such as an oligonucleotide of at least 10, 20, 30, 40, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to mRNA or DNA. The 38911 nucleic acid probe can be, for example, the full-length cDNA of SEQ ID NO:3, or a fragment thereof, such as an oligonucleotide of at least 10, 20, 30, 40, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to mRNA or DNA. The 26904 nucleic acid probe can be, for example, the full-length cDNA of SEQ ID NO:5, or a fragment thereof, such as an oligonucleotide of at least 10, 20, 30, 40, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to mRNA or DNA.

Fragments of the polynucleotides described herein are also useful to synthesize larger fragments or full-length polynucleotides described herein. For example, a fragment can be hybridized to any portion of an mRNA and a larger or full-length cDNA can be produced.

The fragments are also useful to synthesize antisense molecules of desired length and sequence.

Antisense nucleic acids of the invention can be designed using the nucleotide sequences of SEQ ID NOS:2, 4, 6, or 8, and constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules

BNSDOCID: <WO___0149847A2_I_>

or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest.

Additionally, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4:5). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996), *supra*; Perry-O'Keefe *et al.*

BNSDOCID: <WO___0149647A2_I_>

(1996) *Proc. Natl. Acad. Sci. USA 93*:14670. PNAs can be further modified, e.g., to enhance their stability, specificity or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), *supra*, Finn *et al.* (1996) *Nucleic Acids Res. 24*(17):3357-63, Mag *et al.* (1989) *Nucleic Acids Res. 17*:5973, and Peterser *et al.* (1975) *Bioorganic Med. Chem. Lett. 5*:1119.

The nucleic acid molecules and fragments of the invention can also include other appended groups such as peptides (e.g., for targeting host cell proteins *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA 86*:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA 84*:648-652; PCT Publication No. WO 88/0918) or the blood brain barrier (see, e.g., PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol *et al.* (1988) *Bio-Techniques 6*:958-976) or intercalating agents (see, e.g., Zon (1988) *Pharm Res. 5*:539-549).

The polynucleotides of the invention are also useful as primers for PCR to amplify any given region of the polynucleotide.

The polynucleotides are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the polypeptides. Vectors also include insertion vectors, used to integrate into another polynucleotide sequence, such as into the cellular genome, to alter *in situ* expression of the genes and gene products. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The polynucleotides are also useful for expressing antigenic peptides. Peptide regions having a high antigenicity index are shown in Figures 3, 9, and 15.

The polynucleotides are also useful as probes for determining the chromosomal positions of the polynucleotides of the invention by means of *in situ* hybridization methods, such as FISH (for a review of this technique, see Verma *et al.* (1988) *Human Chromosomes: A Manual of Basic Techniques* (Pergamon

BNSDOCID: <WO___0149847A2_I_>

Press, New York), and PCR mapping of somatic cell hybrids. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, *Mendelian Inheritance in Man*, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland *et al.* (1987) *Nature 325*:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a specified gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible form chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

The polynucleotide probes are also useful to determine patterns of the presence of the gene encoding the proteins of the invention and their variants with respect to tissue distribution, for example, whether gene duplication has occurred and whether the duplication occurs in all or only a subset of tissues. The genes

can be naturally occurring or can have been introduced into a cell, tissue, or organism exogenously.

The polynucleotides are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from genes encoding the polynucleotides described herein.

The polynucleotides are also useful for constructing host cells expressing a part, or all, of the polynucleotides and polypeptides of the invention.

The polynucleotides are also useful for constructing transgenic animals expressing all, or a part, of the polynucleotides and polypeptides of the invention.

The polynucleotides are also useful for making vectors that express part, or all, of the polypeptides of the invention.

The polynucleotides are also useful as hybridization probes for determining the level of nucleic acid expression of the nucleic acid molecules of the invention. Accordingly, the probes can be used to detect the presence of, or to determine levels of, the nucleic acid in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the polypeptides described herein can be used to assess gene copy number in a given cell, tissue, or organism. This is particularly relevant in cases in which there has been an amplification of the genes of the invention.

Alternatively, the probe can be used in an *in situ* hybridization context to assess the position of extra copies of the genes of the invention, as on extrachromosomal elements or as integrated into chromosomes in which the gene is not normally found, for example as a homogeneously staining region.

These uses are relevant for diagnosis of disorders involving an increase or decrease in expression relative to normal, such as a proliferative disorder, a differentiative or developmental disorder, or a hematopoietic disorder.

Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant expression or activity of a nucleic acid of the invention, in which a test sample is obtained from a subject and nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of the nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the nucleic acid.

One aspect of the invention relates to diagnostic assays for determining nucleic acid expression as well as activity in the context of a biological sample (e.g., blood, serum, cells, tissue) to determine whether an individual has a disease or disorder, or is at risk of developing a disease or disorder, associated with aberrant nucleic acid expression or activity. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with expression or activity of the nucleic acid molecules.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a protein of the invention, such as by measuring a level of a protein-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a gene of the invention has been mutated.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate nucleic acid expression (e.g., antisense, polypeptides, peptidomimetics, small molecules or other drugs) of the nucleic acid molecules of the invention. A cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of mRNA of the invention in the presence of the candidate compound is compared to the level of expression of the mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. The modulator can bind to the nucleic acid or indirectly modulate expression, such as by interacting with other cellular components that affect nucleic acid expression.

Modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject) in patients or in transgenic animals.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the receptor

BNSDOCID: <WO___0149847A2_I_>

gene. The method typically includes assaying the ability of the compound to modulate the expression of the nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired nucleic acid expression of the nucleic acid molecules of the invention.

The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

Alternatively, candidate compounds can be assayed *in vivo* in patients or in transgenic animals.

The assay for nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway (such as cyclic AMP or phosphatidylinositol turnover). Further, the expression of genes that are up- or down-regulated in response to the protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of mRNA in the presence of the candidate compound is compared to the level of expression of mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

Accordingly, the invention provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate nucleic acid expression of the nucleic acid molecules of the invention. Modulation includes both up-regulation (i.e. activation or agonization)

or down-regulation (suppression or antagonization) or effects on nucleic acid activity (e.g. when nucleic acid is mutated or improperly modified). Treatment is of disorders characterized by aberrant expression or activity of the nucleic acid.

Alternatively, a modulator of nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the nucleic acid expression.

The polynucleotides are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

Monitoring can be, for example, as follows: (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a specified mRNA or genomic DNA of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the mRNA or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the mRNA or genomic DNA in the pre-administration sample with the mRNA or genomic DNA in the post-administration sample or samples; and (vi) increasing or decreasing the administration of the agent to the subject accordingly.

The polynucleotides of the invention are also useful in diagnostic assays for qualitative changes in the nucleic acid, and particularly in qualitative changes that lead to pathology. The polynucleotides can be used to detect mutations in genes of the invention and gene expression products such as mRNA. The polynucleotides can be used as hybridization probes to detect naturally-occurring

BNSDOCID: <WO___0149847A2_I_>

genetic mutations in the gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a protein of the invention.

Mutations in a gene of the invention can be detected at the nucleic acid level by a variety of techniques. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way.

In certain embodiments, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science 241*:1077-1080 (1988); and Nakazawa *et al.*, *PNAS 91*:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res. 23*:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA 87*:1874-1878), transcriptional amplification system (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA 86*:1173-1177), Q-Beta Replicase (Lizardi *et al.* (1988) *Bio/Technology 6*:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well-known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

Alternatively, mutations in a gene of the invention can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method.

Furthermore, sequence differences between a mutant gene of the invention and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques 19*:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al., Adv. Chromatogr. 36*:127-162 (1996); and Griffin *et al., Appl. Biochem. Biotechnol. 38*:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al., Science 230*:1242 (1985)); Cotton *et al., PNAS 85*:4397 (1988); Saleeba *et al., Meth. Enzymol. 217*:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared

(Orita *et al.*, *PNAS 86*:2766 (1989); Cotton *et al.*, *Mutat. Res. 285*:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl. 9*:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature 313*:495 (1985)). The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet. 7*:5). Examples of other techniques for detecting point mutations include, selective oligonucleotide hybridization, selective amplification, and selective primer extension.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotide probes (Cronin *et al.* (1996) *Human Mutation 7*:244-255; Kozal *et al.* (1996) *Nature Medicine 2*:753-759). For example, genetic mutations can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

The polynucleotides of the invention are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the polynucleotides can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). In the present case, for example, a mutation in the gene that results in altered affinity

for ligand could result in an excessive or decreased drug effect with standard concentrations of ligand that activates the protein. Accordingly, the polynucleotides described herein can be used to assess the mutation content of the gene in an individual in order to select an appropriate compound or dosage regimen for treatment.

Thus polynucleotides displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The methods can involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting mRNA, or genomic DNA, such that the presence of mRNA or genomic DNA is detected in the biological sample, and comparing the presence of mRNA or genomic DNA in the control sample with the presence of mRNA or genomic DNA in the test sample.

The polynucleotides are also useful for chromosome identification when the sequence is identified with an individual chromosome and to a particular location on the chromosome. First, the DNA sequence is matched to the chromosome by *in situ* or other chromosome-specific hybridization. Sequences can also be correlated to specific chromosomes by preparing PCR primers that can be used for PCR screening of somatic cell hybrids containing individual chromosomes from the desired species. Only hybrids containing the chromosome containing the gene homologous to the primer will yield an amplified fragment. Sublocalization can be achieved using chromosomal fragments. Other strategies include prescreening with labeled flow-sorted chromosomes and preselection by hybridization to chromosome-specific libraries. Further mapping strategies include fluorescence *in situ* hybridization which allows hybridization with probes shorter than those traditionally used. Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on the chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the

BNSDOCID: <WO___0149847A2_I_>

genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

The polynucleotides can also be used to identify individuals from small biological samples. This can be done for example using restriction fragment-length polymorphism (RFLP) to identify an individual. Thus, the polynucleotides described herein are useful as DNA markers for RFLP (See U.S. Patent No. 5,272,057).

Furthermore, the gene sequence can be used to provide an alternative technique which determines the actual DNA sequence of selected fragments in the genome of an individual. Thus, the receptor sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify DNA from an individual for subsequent sequencing.

Panels of corresponding DNA sequences from individuals prepared in this manner can provide unique individual identifications, as each individual will have a unique set of such DNA sequences. It is estimated that allelic variation in humans occurs with a frequency of about once per each 500 bases. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. The sequences can be used to obtain such identification sequences from individuals and from tissue. The sequences represent unique fragments of the human genome. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes.

If a panel of reagents from the sequences is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

The polynucleotides can also be used in forensic identification procedures. PCR technology can be used to amplify DNA sequences taken from very small biological samples, such as a single hair follicle, body fluids (e.g. blood, saliva, or

BNSDOCID: <WO___0149847A2_I_>

semen). The amplified sequence can then be compared to a standard allowing identification of the origin of the sample.

The polynucleotides can thus be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As described above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to the noncoding region are particularly useful since greater polymorphism occurs in the noncoding regions, making it easier to differentiate individuals using this technique.

The polynucleotides can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This is useful in cases in which a forensic pathologist is presented with a tissue of unknown origin. Panels of probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these primers and probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Alternatively, the polynucleotides can be used directly to block transcription or translation of the gene sequences by means of antisense or ribozyme constructs. Thus, in a disorder characterized by abnormally high or undesirable expression of the gene of the invention, nucleic acids can be directly used for treatment.

The polynucleotides are thus useful as antisense constructs to control expression of a gene of the invention in cells, tissues, and organisms. A DNA antisense polynucleotide is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of the protein of the invention. An antisense RNA or DNA polynucleotide would hybridize to the mRNA and thus block translation of mRNA into protein.

Examples of antisense molecules useful to inhibit nucleic acid expression include antisense molecules complementary to a fragment of the 5' untranslated region of SEQ ID NOS:2, 4, or 6, which also includes the start codon and antisense molecules which are complementary to a fragment of the 3' untranslated region of SEQ ID NOS:2, 4, or 6.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of nucleic acid of the invention. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired expression of a nucleic acid of the invention. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the protein of the invention, such as ligand binding.

The polynucleotides also provide vectors for gene therapy in patients containing cells that are aberrant in expression of a gene of the invention. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired protein to treat the individual.

The invention also encompasses kits for detecting the presence of a nucleic acid of the invention in a biological sample. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting the nucleic acid in a biological sample; means for determining the amount of the nucleic acid in the sample; and means for comparing the amount of the nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect the mRNA or DNA.

Computer Readable Means

The nucleotide or amino acid sequences of the invention are also provided in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid or amino acid molecule,

which contains a nucleotide or amino acid sequence of the present invention. Such a manufacture provides the nucleotide or amino acid sequences, or a subset thereof (e.g., a subset of open reading frames (ORFs)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exists in nature or in purified form.

In one application of this embodiment, a nucleotide or amino acid sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. The skilled artisan will readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. The skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide or amino acid sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The

BNSDOCID: <WO___0149847A2_I_>

skilled artisan can readily adapt any number of dataprocessor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing the nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention.

Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA).

For example, software which implements the BLAST (Altschul *et al.* (1990) *J. Mol. Biol. 215*:403-410) and BLAZE (Brutlag *et al.* (1993) *Comp. Chem. 17*:203-207) search algorithms on a Sybase system can be used to identify open reading frames (ORFs) of the sequences of the invention which contain homology to ORFs or proteins from other libraries. Such ORFs are protein encoding fragments and are useful in producing commercially important proteins such as enzymes used in various reactions and in the production of commercially useful metabolites.

Vectors/host cells

The invention also provides vectors containing the polynucleotides of the invention. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, that can transport the polynucleotides. When the vector is a nucleic acid molecule, the polynucleotides are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the polynucleotides of the invention. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the polynucleotides when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the polynucleotides. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the polynucleotides such that transcription of the polynucleotides is allowed in a host cell. The polynucleotides can be introduced into the host cell with a separate polynucleotide capable of affecting transcription. Thus, the second polynucleotide may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the polynucleotides from

BNSDOCID: <WO___0149847A2_I_>

the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself.

It is understood, however, that in some embodiments, transcription and/or translation of the polynucleotides can occur in a cell-free system.

The regulatory sequence to which the polynucleotides described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ, the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al., Molecular Cloning: A Laboratory Manual, 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a polynucleotide of the invention. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and

BNSDOCID: <WO___0149847A2_I_>

phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The polynucleotides can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate polynucleotide can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the polypeptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the polypeptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired polypeptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene 67*:31-40 (1988)), pMAL (New England Biolabs, Beverly,

BNSDOCID: <WO__0149847A2_I_>

MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al., Gene 69*:301-315 (1988)) and pET 11d (Studier *et al., Gene Expression Technology: Methods in Enzymology 185*:60-89 (1990)).

Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the polynucleotide of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al., Nucleic Acids Res. 20*:2111-2118 (1992)).

The polynucleotides can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al., EMBO J. 6*:229-234 (1987)), pMFa (Kurjan *et al., Cell 30*:933-943 (1982)), pJRY88 (Schultz *et al., Gene 54*:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The polynucleotides can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al., Mol. Cell Biol. 3*:2156-2165 (1983)) and the pVL series (Lucklow *et al., Virology 170*:31-39 (1989)).

In certain embodiments of the invention, the polynucleotides described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature 329*:840 (1987)) and pMT2PC (Kaufman *et al., EMBO J. 6*:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the polynucleotides. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the polynucleotides described herein. These are found for example

in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual, 2nd, ed.*, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the polynucleotide sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual, 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the polynucleotides of the invention can be introduced either alone or with other polynucleotides that are not related to the polynucleotides of the invention such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the polynucleotide vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and

BNSDOCID: <WO___0149847A2_I_>

transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the polynucleotides described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the polypeptide is desired, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the polypeptides of the invention or heterologous to these polypeptides.

Where the polypeptide is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The polypeptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the polypeptides described herein, the polypeptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the polypeptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

It is understood that "host cells" and "recombinant host cells" refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

The host cells expressing the polypeptides described herein, and particularly recombinant host cells, have a variety of uses. First, the cells are useful for producing proteins or polypeptides of the invention that can be further purified to produce desired amounts of the protein or fragments. Thus, host cells containing expression vectors are useful for polypeptide production.

Host cells are also useful for conducting cell-based assays involving the protein or fragments. Thus, a recombinant host cell expressing a native protein of the invention is useful to assay for compounds that stimulate or inhibit protein function. This includes ligand binding, gene expression at the level of transcription or translation, G-protein interaction, and components of the signal transduction pathway.

Host cells are also useful for identifying mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native protein.

Recombinant host cells are also useful for expressing the chimeric polypeptides described herein to assess compounds that activate or suppress activation by means of a heterologous amino terminal extracellular domain (or other binding region). Alternatively, a heterologous region spanning the entire transmembrane domain (or parts thereof) can be used to assess the effect of a desired amino terminal extracellular domain (or other binding region) on any given host cell. In this embodiment, a region spanning the entire transmembrane

BNSDOCID: <WO___0149847A2_I_>

domain (or parts thereof) compatible with the specific host cell is used to make the chimeric vector. Alternatively, a heterologous carboxy terminal intracellular, e.g., signal transduction, domain can be introduced into the host cell.

Further, mutant proteins can be designed in which one or more of the various functions is engineered to be increased or decreased (e.g., ligand binding or G-protein binding) and used to augment or replace the native proteins in an individual. Thus, host cells can provide a therapeutic benefit by replacing an aberrant protein of the invention or providing an aberrant protein that provides a therapeutic result. In one embodiment, the cells provide proteins that are abnormally active.

In another embodiment, the cells provide proteins that are abnormally inactive. These proteins can compete with the endogenous proteins in the individual.

In another embodiment, cells expressing proteins that cannot be activated, are introduced into an individual in order to compete with the endogenous proteins for ligand. For example, in the case in which excessive ligand is part of a treatment modality, it may be necessary to inactivate this ligand at a specific point in treatment. Providing cells that compete for the ligand, but which cannot be affected by receptor activation would be beneficial.

Homologously recombinant host cells can also be produced that allow the *in situ* alteration of the endogenous polynucleotide sequences in a host cell genome. The host cell includes, but is not limited to, a stable cell line, cell *in vivo*, or cloned microorganism. This technology is more fully described in WO 93/09222, WO 91/12650, WO 91/06667, U.S. 5,272,071, and U.S. 5,641,670. Briefly, specific polynucleotide sequences corresponding to the polynucleotides or sequences proximal or distal to a gene of the invention are allowed to integrate into a host cell genome by homologous recombination where expression of the gene can be affected. In one embodiment, regulatory sequences are introduced that either increase or decrease expression of an endogenous sequence. Accordingly, a protein of the invention can be produced in a cell not normally producing it. Alternatively, increased expression of the protein can be effected in a cell normally producing the protein at a specific level. Further, expression can

be decreased or eliminated by introducing a specific regulatory sequence. The regulatory sequence can be heterologous to the protein sequence or can be a homologous sequence with a desired mutation that affects expression. Alternatively, the entire gene can be deleted. The regulatory sequence can be specific to the host cell or capable of functioning in more than one cell type. Still further, specific mutations can be introduced into any desired region of the gene to produce mutant proteins. Such mutations could be introduced, for example, into the specific functional regions such as the ligand-binding site.

In one embodiment, the host cell can be a fertilized oocyte or embryonic stem cell that can be used to produce a transgenic animal containing the altered gene. Alternatively, the host cell can be a stem cell or other early tissue precursor that gives rise to a specific subset of cells and can be used to produce transgenic tissues in an animal. See also Thomas *et al., Cell 51*:503 (1987) for a description of homologous recombination vectors. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous receptor gene is selected (see e.g., Li, E. *et al., Cell 69*:915 (1992)). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A.. (1991) *Current Opinions in Biotechnology 2*:823-829 and in PCT International Publication Nos. WO 90/11354; WO 91/01140; and WO 93/04169.

The genetically engineered host cells can be used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the

BNSDOCID: <WO___0149847A2_I_>

genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a receptor protein and identifying and evaluating modulators of the protein activity.

Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

In one embodiment, a host cell is a fertilized oocyte or an embryonic stem cell into which the polynucleotide sequences have been introduced.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the nucleotide sequences of the invention can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS 89*:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science 251*:1351-1355 (1991)). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature 385*:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the polypeptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, receptor activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* receptor function, including ligand interaction, the effect of specific mutant receptors on receptor function and ligand interaction, and the effect of chimeric

BNSDOCID: <WO___0149847A2_I_>

receptors. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more receptor functions.

In general, methods for producing transgenic animals include introducing a nucleic acid sequence according to the present invention, the nucleic acid sequence capable of expressing the protein in a transgenic animal, into a cell in culture or *in vivo*. When introduced *in vivo*, the nucleic acid is introduced into an intact organism such that one or more cell types and, accordingly, one or more tissue types, express the nucleic acid encoding the protein. Alternatively, the nucleic acid can be introduced into virtually all cells in an organism by transfecting a cell in culture, such as an embryonic stem cell, as described herein for the production of transgenic animals, and this cell can be used to produce an entire transgenic organism. As described, in a further embodiment, the host cell can be a fertilized oocyte. Such cells are then allowed to develop in a female foster animal to produce the transgenic organism.

Pharmaceutical compositions

The nucleic acid molecules of the invention, protein of the invention (particularly fragments such as the amino terminal extracellular domain), modulators of the protein, and antibodies (also referred to herein as "active compounds") can be incorporated into pharmaceutical compositions suitable for administration to a subject, e.g., a human. Such compositions typically comprise the nucleic acid molecule, protein, modulator, or antibody and a pharmaceutically acceptable carrier.

As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, such media can be used in the compositions of the invention. Supplementary active compounds can also be incorporated into the compositions. A pharmaceutical composition of the invention is formulated to be compatible with its intended route

of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the

BNSDOCID: <WO__0149847A2_I_>

injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a seven-transmembrane protein/receptor protein or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For oral administration, the agent can be contained in enteric forms to survive the stomach or further coated or mixed to be released in a particular region of the GI tract by known methods. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the

BNSDOCID: <WO__0149847A2_I_>

dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. 5,328,470) or by stereotactic injection (see e.g., Chen *et al., PNAS 91*:3054-3057 (1994)). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight.

The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments. In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or

polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low

BNSDOCID: <WO___0149847A2_I_>

dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

This invention may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will fully convey the invention to those skilled in the art. Many modifications and other embodiments of the invention will come to mind in one skilled in the art to which this invention pertains having the benefit of the teachings presented in the foregoing description. Although specific terms are employed, they are used as in the art unless otherwise indicated.

Applicant's or agent's file reference	35800/207180	International application No.

INDICATIONS RELATING TO DEPOSITED MICROORGANISM OR OTHER BIOLOGICAL MATERIAL

(PCT Rule 13*bis*)

A. The indications made below relate to the deposited microorganism or other biological material referred to in the description on page 7, line 16

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depository institution
American Type Culture Collection

Address of depositary institution *(including postal code and country)*
10801 University Blvd.
Manassas, VA 20110-2209 US

Date of deposit	Accession Number
06 April 2000 (06.04.00)	PTA-1654

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

Page 17, line 6; page 114, lines 8,12,16,20,24,28; page 115, lines 5,10,23,27,31; page 116, lines 4,8,12,25, 29; page 117, line 2,6,10,14,23,26,30; page 118, lines 11,17,27; page 119, lines 9,12,16,29,32; page 120, lines 4,22,25,29; page 121, lines 20,24,28,32; page 122, lines 4,8,26,30; page 123, lines 2,6,10,14,24,27,31; page 124, lines 23,26,30; page 125, lines 15,18,22; page 126, lines 6,10,14,18,22,26; page 127, lines 9,13,17,21,25,29; page 128, lines 11,14,18

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indicators are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received with the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1998)

BNSDOCID: <WO __0149847A2_I_>

Applicant's or agent's file reference 35800/207180	International application No.

INDICATIONS RELATING TO DEPOSITED MICROORGANISM OR OTHER BIOLOGICAL MATERIAL

(PCT Rule 13*bis*)

A. The indications made below relate to the deposited microorganism or other biological material referred to in the description on page 7, line 12

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depository institution
American Type Culture Collection

Address of depositary institution *(including postal code and country)*
10801 University Blvd.
Manassas, VA 20110-2209 US

Date of deposit	Accession Number
09 May 2000 (09.05.00)	PTA-1847

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

Page 17, line 2; page 114, lines 8,12,16,21,24,29; page 115, lines 5,10,23,27,31; page 116, lines 5,8,13,25, 29; page 117, line 2,7,10,15,23,26,30; page 118, lines 11,16,22; page 119, lines 9,12,16,29,32; page 120, lines 4,22,25,29; page 121, lines 20,24,28; page 122, lines 1,4,9,26,30; page 123, lines 2,7,10,15,24,27,31; page 124, lines 23,26,30; page 125, lines 15,18,22; page 126, lines 6,10,14,19,22,27; page 127, lines 9,13,17,22,25,30; page 128, lines 11,14,18

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indicators are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received with the International Bureau on:
Authorized officer	Authorized officer

THAT WHICH IS CLAIMED:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

b) a nucleic acid molecule having a fragment of at least 15 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2, 4, or 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

c) a nucleic acid molecule encoding a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

d) a nucleic acid molecule encoding a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847; and

e) a nucleic acid molecule encoding a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule having SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions.

BNSDOCID: <WO___0149847A2_I_>

2. The isolated nucleic acid molecule of claim 1, wherein said nucleic acid molecule is selected from the group consisting of:

a) a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2, 4, or 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof; and

b) a nucleic acid molecule encoding a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847.

3. The nucleic acid molecule of claim 1, further having vector nucleotide sequences.

4. The nucleic acid molecule of claim 1, further having a nucleotide sequence encoding at least one heterologous polypeptide.

5. A host cell that contains a nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

b) a nucleic acid molecule having a fragment of at least 15 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2, 4, or 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

c) a nucleic acid molecule encoding a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

d) a nucleic acid molecule encoding a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847; and

e) a nucleic acid molecule encoding a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule having SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions.

6. The host cell of claim 5, wherein said host cell is a mammalian host cell.

7. A nonhuman mammalian host cell containing at least one nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

b) a nucleic acid molecule having a fragment of at least 15 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2, 4, or 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

c) a nucleic acid molecule encoding a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence

encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

d) a nucleic acid molecule encoding a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847; and

e) a nucleic acid molecule encoding a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule having SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions.

8. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

b) a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the polypeptide is encoded by a nucleic acid molecule that hybridizes to a

nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions; and

c) a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, or 6, or a complement thereof.

9. The isolated polypeptide of claim 8, wherein said polypeptide has an amino acid sequence selected from the group consisting of:

(a) SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847.

(b) The amino acid sequence set forth as about amino acid 6 to about amino acid 337 of SEQ ID NO:1 or SEQ ID NO:3;

(c) The amino acid sequence extending from about amino acid 6 to about amino acid 337 of the polypeptide encoded by the cDNA insert of the plasmid deposited with ATCC as Patent Deposit No. PTA-1847 or PTA-1654;

(d) The amino acid sequence set forth as about amino acid 1 to about amino acid 37 of SEQ ID NO:1;

(e) The amino acid sequence extending from about amino acid 1 to about amino acid 37 of the polypeptide encoded by the cDNA contained in ATCC Deposit No. PTA-1847;

(f) The amino acid sequence set forth as about amino acid 1 to about amino acid 40 of SEQ ID NO:3;

(g) The amino acid sequence extending from about amino acid 1 to about amino acid 40 of the polypeptide encoded by the cDNA insert of the plasmid deposited with ATCC as Patent Deposit No. PTA-1654; and

(h) The amino acid set forth as about amino acid 6 to about amino acid 450 of SEQ ID NO:5.

BNSDOCID: <WO 0149847A2_I_>

10. The polypeptide of claim 8, further having heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide, wherein said polypeptide is selected from the group consisting of:

a) a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

b) a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions; and

c) a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, or 6, or a complement thereof.

12. A method for producing a polypeptide selected from the group consisting of:

a) a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

b) a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions; and

c) a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, or 6, or a complement thereof;

said method comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

13. The method of claim 12, wherein said polypeptide has the amino acid sequence of SEQ ID NO:1, 3, or 5.

14. A method for detecting the presence of a polypeptide in a sample, wherein said polypeptide is selected from the group consisting of:

a) a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

b) a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions; and

c) a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, or 6, or a complement thereof;

said method having the steps of contacting the sample with a compound that selectively binds to the polypeptide and determining whether the compound binds to the polypeptide in the sample.

15. The method of claim 14, wherein the compound that binds to the polypeptide is an antibody.

16. A kit having a compound that selectively binds to a polypeptide of claim 8 and instructions for use.

17. A method for detecting the presence of a nucleic acid molecule in a sample, wherein said nucleic acid molecule is selected from the group consisting of:

a) a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

b) a nucleic acid molecule having a fragment of at least 15 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2, 4, or 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

c) a nucleic acid molecule encoding a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

d) a nucleic acid molecule encoding a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or

PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847; and

e) a nucleic acid molecule encoding a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule having SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions;

said method having the steps of contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

18. The method of claim 17, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

19. A kit having a compound which selectively hybridizes to a nucleic acid molecule and instructions for use, wherein the nucleic acid molecule is selected from the group consisting of:

a) a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

b) a nucleic acid molecule having a fragment of at least 15 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2, 4, or 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

c) a nucleic acid molecule encoding a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence

RNSDOCID: <WO 0149847A2 I_>

encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

d) a nucleic acid molecule encoding a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847; and

e) a nucleic acid molecule encoding a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule having SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions.

20. A method for identifying a compound which binds to a polypeptide, wherein said polypeptide is selected from the group consisting of:

a) a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

b) a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the polypeptide is encoded by a nucleic acid molecule that hybridizes to a

nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions; and

c) a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, or 6, or a complement thereof;

said method having the steps of contacting the polypeptide or a cell expressing the polypeptide with a test compound and determining whether the polypeptide binds to the test compound.

21. The method of claim 20, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detection of test compound/polypeptide binding;

b) detection of binding using a competition binding assay;

c) detection of binding using an assay for GPCR-ligand binding.

22. A method for modulating the activity of a polypeptide selected from the group consisting of:

a) a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

b) a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the polypeptide is encoded by a nucleic acid molecule that hybridizes to a

nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions; and

c) a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, or 6, or a complement thereof;

said method having the steps of contacting the polypeptide or a cell expressing the polypeptide with a compound that binds to the polypeptide, under conditions in which the compound is capable of modulating the activity of the polypeptide.

23. A method for identifying a compound which modulates the activity of a polypeptide selected from the group consisting of:

a) a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

b) a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions; and

c) a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, or 6, or a complement thereof;

said method having the steps of contacting the polypeptide with a test compound and determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

24. A method for identifying an agent that modulates the level of expression of a nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

b) a nucleic acid molecule having a fragment of at least 15 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2, 4, or 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

c) a nucleic acid molecule encoding a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

d) a nucleic acid molecule encoding a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847; and

e) a nucleic acid molecule encoding a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule having SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions;

said method having the steps of contacting said agent with a cell expressing said nucleic acid molecule, under conditions such that said level of expression of said

BNSDOCID: <WO 0149847A2 I >

nucleic acid molecule can be modulated in said cell by said agent; and measuring the level of expression of said nucleic acid molecule.

25. A method for modulating the level of expression of a nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

b) a nucleic acid molecule having a fragment of at least 15 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2, 4, or 6, the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, or a complement thereof;

c) a nucleic acid molecule encoding a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

d) a nucleic acid molecule encoding a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847; and

e) a nucleic acid molecule encoding a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule having SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions;

said method comprising contacting said nucleic acid molecule with an agent under conditions that allow the agent to modulate the level of expression of the nucleic acid molecule.

26. A pharmaceutical composition containing at least one polypeptide in a pharmaceutically acceptable carrier, wherein said polypeptide is selected from the group consisting of:

a) a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the fragment has at least 12 contiguous amino acids of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847;

b) a naturally occurring allelic variant of a polypeptide having the amino acid sequence of SEQ ID NO:1, 3, or 5, or an amino acid sequence encoded by the cDNA insert of any of the plasmids deposited with ATCC as Patent Deposit Number PTA-1654 or PTA-1847, wherein the polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleotide sequence having the nucleotide sequence set forth in SEQ ID NO:2, 4, or 6, or a complement thereof under stringent conditions; and

c) a polypeptide encoded by a nucleic acid molecule having a nucleotide sequence that is at least 60% identical to the nucleotide sequence of SEQ ID NO:2, 4, or 6, or a complement thereof.

BNSDOCID: <WO___0149847A2_I_>

Input file Fbh39404FL.seq; Output File 39404.trans
Sequence length 1729

```
CCTTTTTTTTTTTTTTTTTAACTTTTATATTTTTATTAGATGCATTTAGTAACTTGCCTCATAGTCATTTTCTTGGAAAT

TCAATTTCTTCTCCACAGGGTCTCTTTTGAGATTAAAGAGAGAGAAGTGGCAAATTTAGGATGTTAGAATAATTTTCAT

TTAAAAGTAGATCCTTGTTTTTATTACCCTATCATTAATGTTTTCTGTTTTCCTTTATCAGCGAGTTACTGCTCATTTG

                                                    M   N   E   P   L          5
ATTCATATTGCCAAACTGAACTCTCTTGTTTTCTTGCAAGATGAAAGGAGACAACC ATG AAT GAG CCA CTA    15

 D   Y   L   A   N   A   S   D   F   P   D   Y   A   A   A   F   G   N   C   T   25
GAC TAT TTA GCA AAT GCT TCT GAT TTC CCC GAT TAT GCA GCT GCT TTT GGA AAT TGC ACT   75

 D   E   N   I   P   L   K   M   H   Y   L   P   V   I   Y   G   I   I   F   L   45
GAT GAA AAC ATC CCA CTC AAG ATG CAC TAC CTC CCT GTT ATT TAT GGC ATT ATC TTC CTC  135

 V   G   F   P   G   N   A   V   V   I   S   T   Y   I   F   K   M   R   P   W   65
GTG GGA TTT CCA GGC AAT GCA GTA GTG ATA TCC ACT TAC ATT TTC AAA ATG AGA CCT TGG  195

 K   S   S   T   I   I   M   L   N   L   A   C   T   D   L   L   Y   L   T   S   85
AAG AGC AGC ACC ATC ATT ATG CTG AAC CTG GCC TGC ACA GAT CTG CTG TAT CTG ACC AGC  255

 L   P   F   L   I   H   Y   Y   A   S   G   E   N   W   I   F   G   D   F   M  105
CTC CCC TTC CTG ATT CAC TAC TAT GCC AGT GGC GAA AAC TGG ATC TTT GGA GAT TTC ATG  315

 C   K   F   I   R   F   S   F   H   F   N   L   Y   S   S   I   L   F   L   T  125
TGT AAG TTT ATC CGC TTC AGC TTC CAT TTC AAC CTG TAT AGC AGC ATC CTC TTC CTC ACC  375

 C   F   S   I   F   R   Y   C   V   I   I   H   P   M   S   C   F   S   I   H  145
TGT TTC AGC ATC TTC CGC TAC TGT GTG ATC ATT CAC CCA ATG AGC TGC TTT TCC ATT CAC  435

 K   T   R   C   A   V   V   A   C   A   V   V   W   I   I   S   L   V   A   V  165
AAA ACT CGA TGT GCA GTT GTA GCC TGT GCT GTG GTG TGG ATC ATT TCA CTG GTA GCT GTC  495

 I   P   M   T   F   L   I   T   S   T   N   R   T   N   R   S   A   C   L   D  185
ATT CCG ATG ACC TTC TTG ATC ACA TCA ACC AAC AGG ACC AAC AGA TCA GCC TGT CTC GAC  555

 L   T   S   S   D   E   L   N   T   I   K   W   Y   N   L   I   L   T   A   T  205
CTC ACC AGT TCG GAT GAA CTC AAT ACT ATT AAG TGG TAC AAC CTG ATT TTG ACT GCA ACT  615

 T   F   C   L   P   L   V   I   V   T   L   C   Y   T   T   I   I   H   T   L  225
ACT TTC TGC CTC CCC TTG GTG ATA GTG ACA CTT TGC TAT ACC ACG ATT ATC CAC ACT CTG  675

 T   H   G   L   Q   T   D   S   C   L   K   Q   K   A   R   R   L   T   I   L  245
ACC CAT GGA CTG CAA ACT GAC AGC TGC CTT AAG CAG AAA GCA CGA AGG CTA ACC ATT CTG  735

 L   L   L   A   F   Y   V   C   F   L   P   F   H   I   L   R   V   I   R   I  265
CTA CTC CTT GCA TTT TAC GTA TGT TTT TTA CCC TTC CAT ATC TTG AGG GTC ATT CGG ATC  795

 E   S   R   L   L   S   I   S   C   S   I   E   N   Q   I   H   E   A   Y   I  285
GAA TCT CGC CTG CTT TCA ATC AGT TGT TCC ATT GAG AAT CAG ATC CAT GAA GCT TAC ATC  855

 V   S   G   P   L   A   A   L   N   T   F   G   N   L   L   L   Y   V   V   V  305
GTT TCT GGA CCA TTA GCT GCT CTG AAC ACC TTT GGT AAC CTG TTA CTA TAT GTG GTG GTC  915

 S   D   N   F   Q   Q   A   V   C   S   T   V   R   C   K   V   S   G   N   L  325
AGC GAC AAC TTT CAG CAG GCT GTC TGC TCA ACA GTG AGA TGC AAA GTA AGC GGG AAC CTT  975

 E   Q   A   K   K   I   S   Y   S   N   N   P   *                              338
```

FIG 1

GAG CAA GCA AAG AAA ATT AGT [illegible] AAC AAC CCT TGA

AATATTTCATTTACTTAACCAAAAACAAATACTTGCTGATACTTTACCTAGCATCCTAAGATGTTCAGGATGTCTCCCT

CAATGGAACTCCTGGTAAATACTGTGTATTCAAGTAATCATGTGCCAAAGCCAGGGCAGAGCTTCTAGTTCTTTGCAAT

CCCTTTATTGAGCTCCTCCACTGGGGAGATATAAGAATGGGATGCATGTATATCAGCAAAGTATTCAGACATAGTATTA

CAAGCTATTGGAACTCAGAGGCATCTTAGAGAACATCTGTTCCCACCAACTTACTATATATACACGGAAACCAATTTCT

TACCCTTGCCCTAGATTGCTCAGTAAATTTGTGCCAAGATAGGAGAAAACCAATCTTTTCACTCATCATTTCATGCTTC

TCTGCACTCTGGGCCTATTTGTATTGA

FIG 1 (CONT)

BNSDOCID: <WO 0149847A2 | >

Analysis of 39404 (337 aa)

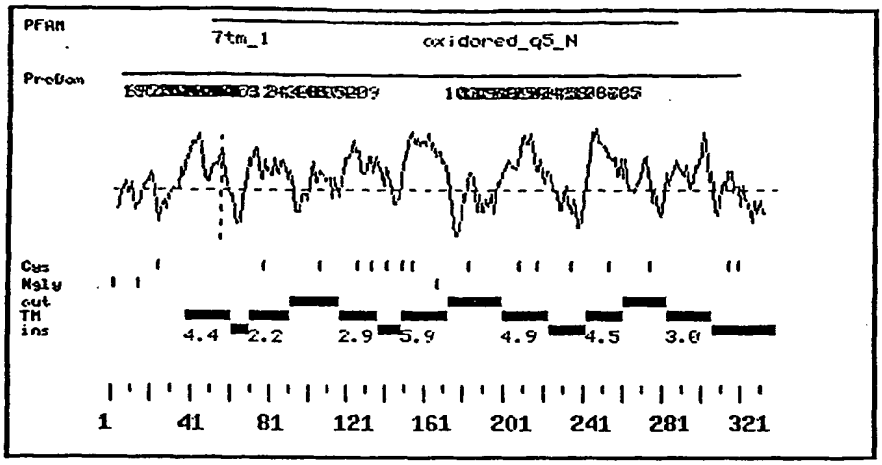


FIG 2

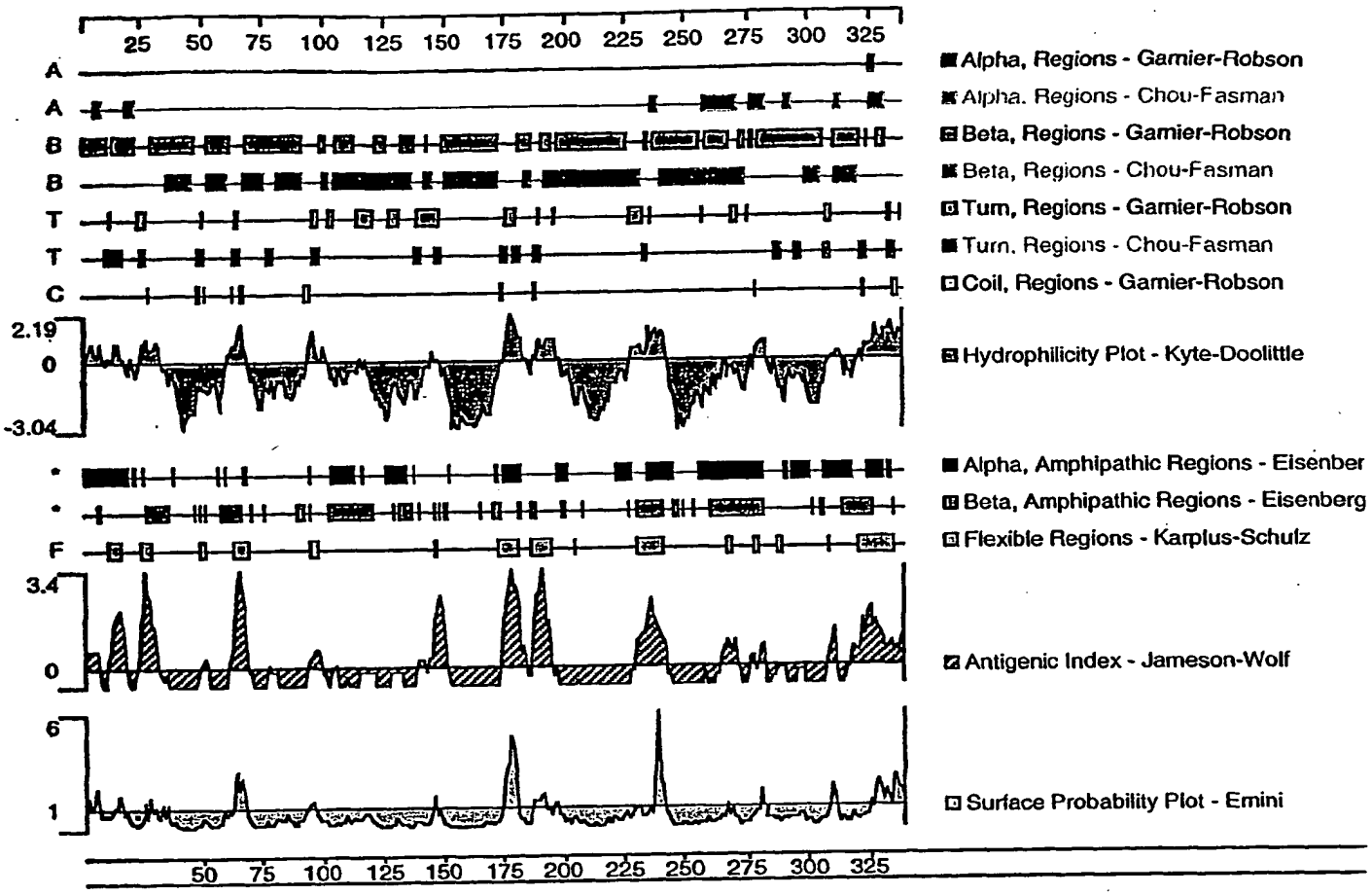

25 50 75 100 125 150 175 200 225 250 275 300 325
A
A
B
B
T
T
C
2.19
0
-3.04
Alpha, Regions - Garnier-Robson
Alpha, Regions - Chou-Fasman
Beta, Regions - Garnier-Robson
Beta, Regions - Chou-Fasman
Turn, Regions - Garnier-Robson
Turn, Regions - Chou-Fasman
Coil, Regions - Garnier-Robson
Hydrophilicity Plot - Kyte-Doolittle
Alpha, Amphipathic Regions - Eisenber
Beta, Amphipathic Regions - Eisenberg
Flexible Regions - Karplus-Schulz
3.4
0
Antigenic Index - Jameson-Wolf
6
1
Surface Probability Plot - Emini
50 75 100 125 150 175 200 225 250 275 300 325

FIG 3

Method	Predict	Score	t@
SignalP (eukaryote)	MAYBE		

Signal Peptide Predictions td. 39404

Note: amino-terminal 70aa used for signal peptide prediction

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
38	60	out-->ins	4.4
70	90	ins-->out	2.2
117	136	out-->ins	2.9
149	172	ins-->out	5.9
200	222	out-->ins	4.9
242	260	ins-->out	4.5
283	305	out-->ins	3.0

Start	End	Orient	Score
16	36	ins-->out	2.2
63	82	out-->ins	2.9
95	118	ins-->out	5.9
146	168	out-->ins	4.9
188	206	ins-->out	4.5
229	251	out-->ins	3.0

Prosite Pattern Matches for 39404

Prosite version: Release 12.2 of February 1995

>PS00001|PDOC00001|ASN_GLYCOSYLATION N-glycosylation site.

Query: 10 NASD 13

Query: 23 NCTD 26

Query: 176 NRTN 179

>PS00004|PDOC00004|CAMP_PHOSPHO_SITE cAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 240 RRLT 243

Query: 329 KKIS 332

>PS00005|PDOC00005|PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 175 TNR 177

Query: 178 TNR 180

Query: 194 TIK 196

Query: 316 TVR 318

>PS00006|PDOC00006|CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query: 187 TSSD 190

FIG 4

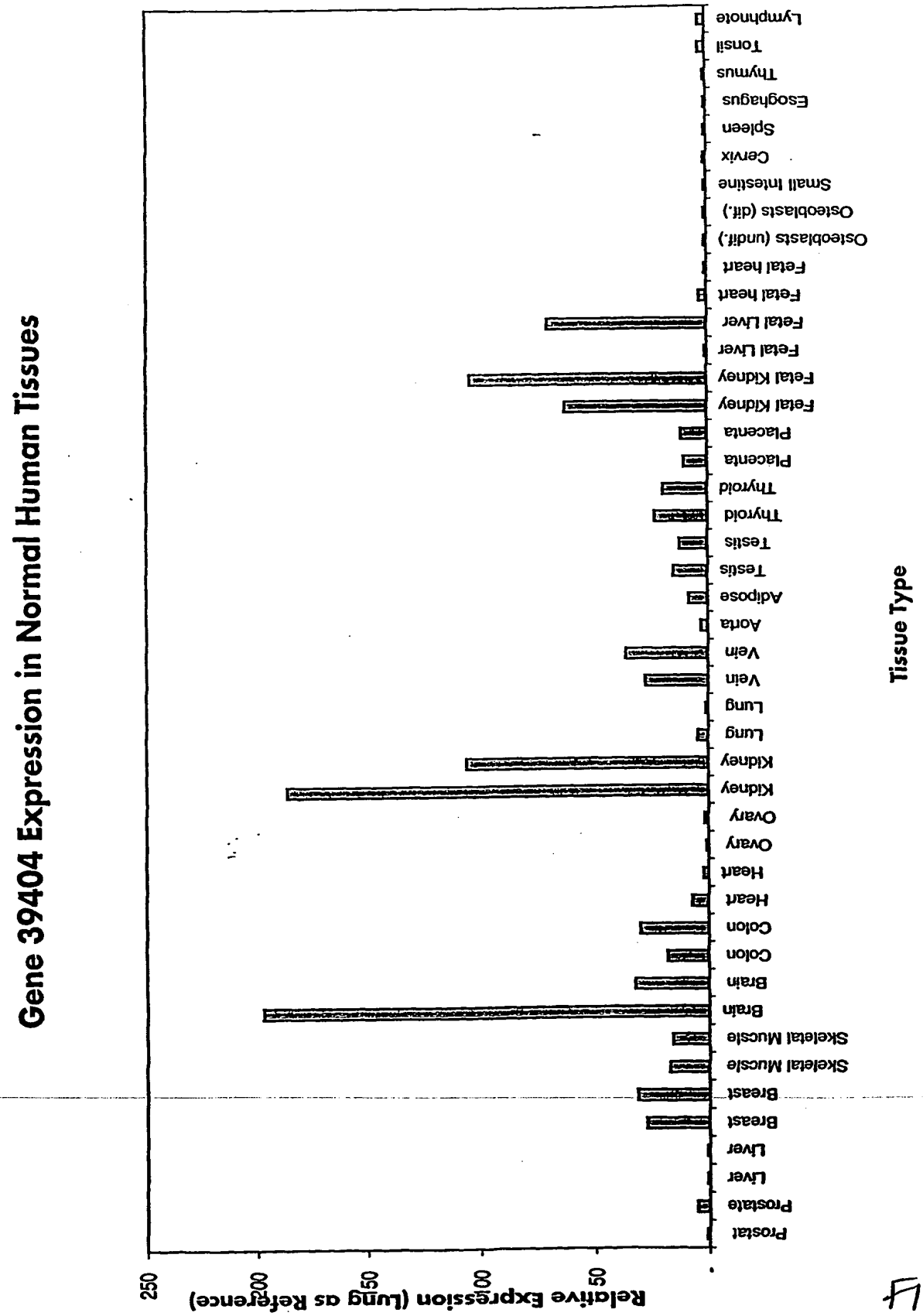

Gene 39404 Expression in Normal Human Tissues
Relative Expression (Lung as Reference)
250
200
150
100
50
Tissue Type
Prostat
Prostate
Liver
Liver
Breast
Breast
Skeletal Muscle
Skeletal Muscle
Brain
Brain
Colon
Colon
Heart
Heart
Ovary
Ovary
Kidney
Kidney
Lung
Lung
Vein
Vein
Aorta
Adipose
Testis
Testis
Thyroid
Thyroid
Placenta
Placenta
Fetal Kidney
Fetal Kidney
Fetal Liver
Fetal Liver
Fetal Heart
Fetal Heart
Osteoblasts (undif.)
Osteoblasts (dif.)
Small Intestine
Cervix
Spleen
Esophagus
Thymus
Tonsil
Lymphnote

FIG 5

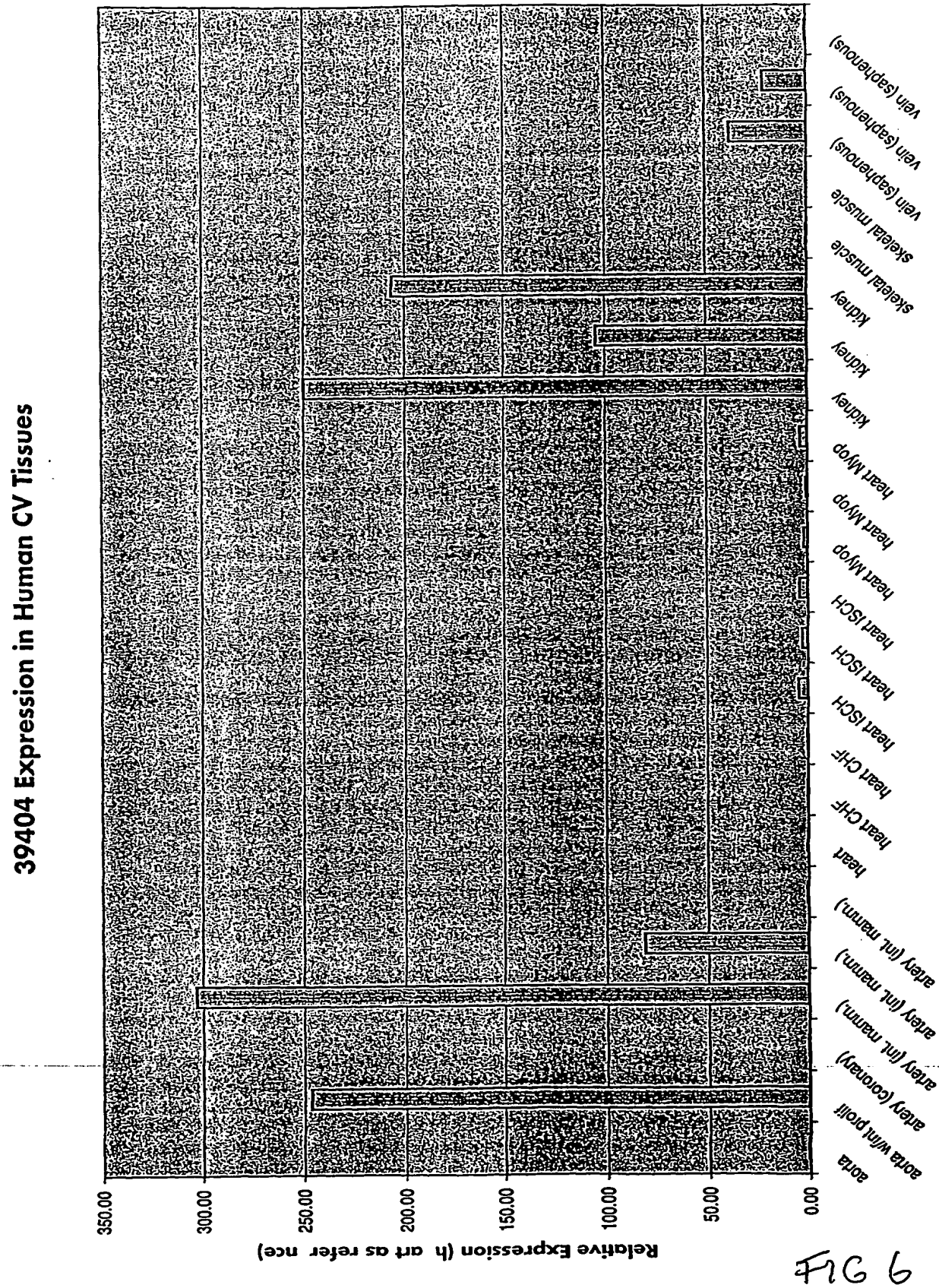

39404 Expression in Human CV Tissues
Relative Expression (h art as refer nce)
350.00
300.00
250.00
200.00
150.00
100.00
50.00
0.00
aorta
aorta w/int probl
artery (coronary)
artery (int. mamm.)
artery (int. mamm.)
artery (int. mamm.)
heart
heart CHF
heart CHF
heart ISCH
heart ISCH
heart ISCH
heart Myop
heart Myop
heart Myop
kidney
kidney
kidney
skeletal muscle
skeletal muscle
vein (saphenous)
vein (saphenous)
vein (saphenous)

FIG 6

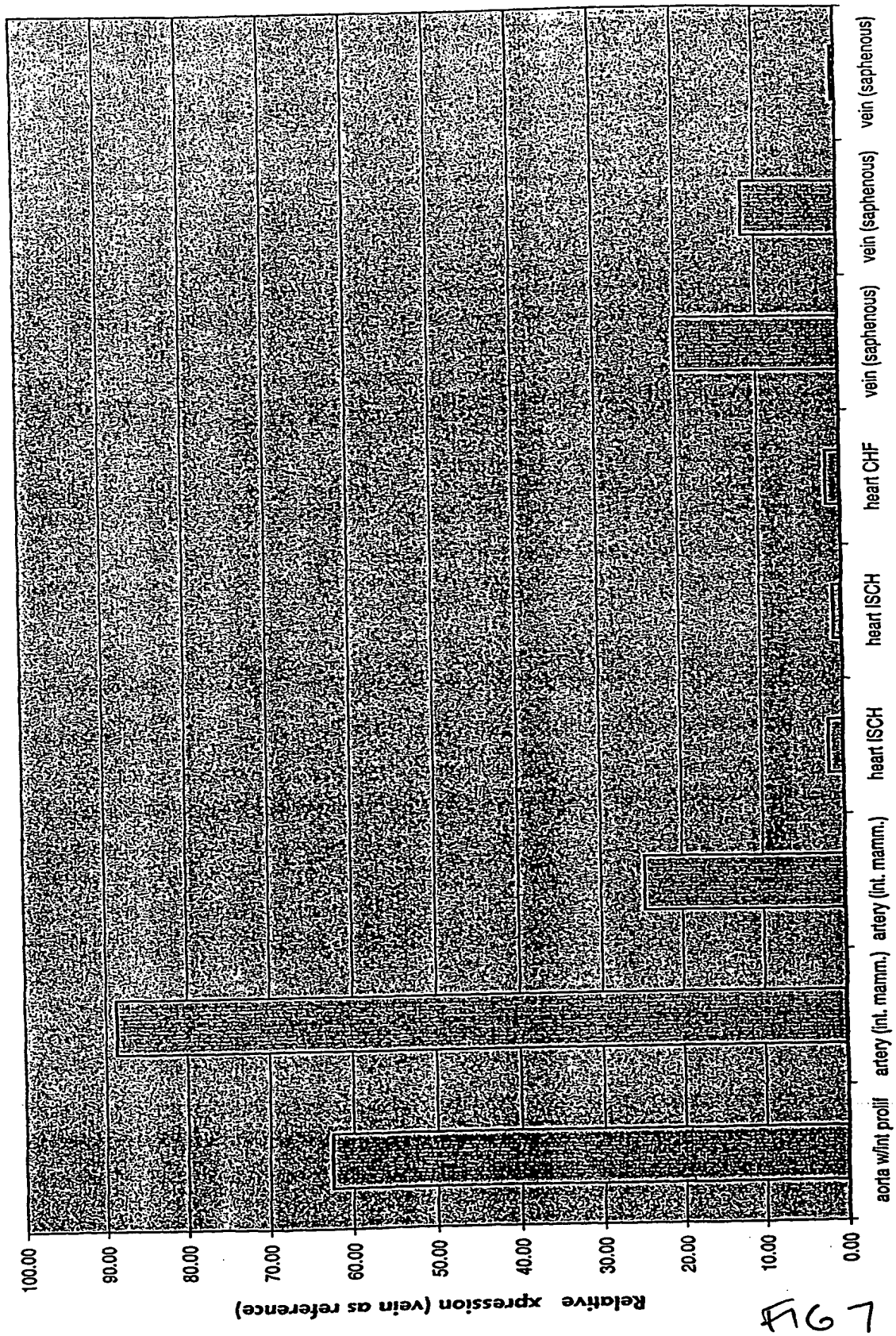

39404 Expression in Human CV tissues
Relative xpression (vein as reference)
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
aorta w/int prolif
artery (int. mamm.)
artery (int. mamm.)
heart ISCH
heart ISCH
heart CHF
vein (saphenous)
vein (saphenous)
vein (saphenous)

FIG 7

Input file Fbh38911a.seq; Output File 38911.trans
Sequence length 1334

```
                10        20        30        40        50        60          M   G   N       3
GTCCGACGTGCTGGACAAATCTTAACTCCTCAAGGACTCCCAAAACCAGAGACACCAGGAGCCTGA ATG GGG AAC       9   75
                                                                                 67

 D   S   V   S   Y   E   Y   G   D   Y   S   D   L   S   D   R   P   V   D   C      23
GAT TCT GTC AGC TAC GAG TAT GGG GAT TAC AGC GAC CTC TCG GAC CGC CCT GTG GAC TGC      69

 L   D   G   A   C   L   A   I   D   P   L   R   V   A   P   L   P   L   Y   A      43
CTG GAT GGC GCC TGC CTG GCC ATC GAC CCG CTG CGC GTG GCC CCG CTC CCA CTG TAT GCC     129

 A   I   F   L   V   G   V   P   G   N   A   M   V   A   W   V   A   G   K   V      63
GCC ATC TTC CTG GTG GGG GTG CCG GGC AAT GCC ATG GTG GCC TGG GTG GCT GGG AAG GTG     189

 A   R   R   R   V   G   A   T   W   L   L   H   L   A   V   A   D   L   L   C      83
GCC CGC CGG AGG GTG GGT GCC ACC TGG TTG CTC CAC CTG GCC GTG GCG GAT TTG CTG TGC     249

 C   L   S   L   P   I   L   A   V   P   I   A   R   G   G   H   W   P   Y   G     103
TGT TTG TCT CTG CCC ATC CTG GCA GTG CCC ATT GCC CGT GGA GGC CAC TGG CCG TAT GGT     309

 A   V   G   C   R   A   L   P   S   I   I   L   L   T   M   Y   A   S   V   L     123
GCA GTG GGC TGT CGG GCG CTG CCC TCC ATC ATC CTG CTG ACC ATG TAT GCC AGC GTC CTG     369

 L   L   A   A   L   S   A   D   L   C   F   L   A   L   G   P   A   W   W   S     143
CTC CTG GCA GCT CTC AGT GCC GAC CTC TGC TTC CTG GCT CTC GGG CCT GCC TGG TGG TCT     429

 T   V   Q   R   A   C   G   V   Q   V   A   C   G   A   A   W   T   L   A   L     163
ACG GTT CAG CGG GCG TGC GGG GTG CAG GTG GCC TGT GGG GCA GCC TGG ACA CTG GCC TTG     489

 L   L   T   V   P   S   A   I   Y   R   R   L   H   Q   E   H   F   P   A   R     183
CTG CTC ACC GTG CCC TCC GCC ATC TAC CGC CGG CTG CAC CAG GAG CAC TTC CCA GCC CGG     549

 L   Q   C   V   V   D   Y   G   G   S   S   S   T   E   N   A   V   T   A   I     203
CTG CAG TGT GTG GTG GAC TAC GGC GGC TCC TCC AGC ACC GAG AAT GCG GTG ACT GCC ATC     609

 R   F   L   F   G   F   L   G   P   L   V   A   V   A   S   C   H   S   A   L     223
CGG TTT CTT TTT GGC TTC CTG GGG CCC CTG GTG GCC GTG GCC AGC TGC CAC AGT GCC CTC     669

 L   C   W   A   A   R   R   C   R   P   L   G   T   A   I   V   V   G   F   F     243
CTG TGC TGG GCA GCC CGA CGC TGC CGG CCG CTG GGC ACA GCC ATT GTG GTG GGG TTT TTT     729

 V   C   W   A   P   Y   H   L   L   G   L   V   L   T   V   A   A   P   N   S     263
GTC TGC TGG GCA CCC TAC CAC CTG CTG GGG CTG GTG CTC ACT GTG GCG GCC CCG AAC TCC     789

 A   L   L   A   R   A   L   R   A   E   P   L   I   V   G   L   A   L   A   H     283
GCA CTC CTG GCC AGG GCC CTG CGG GCT GAA CCC CTC ATC GTG GGC CTT GCC CTC GCT CAC     849

 S   C   L   N   P   M   L   F   L   Y   F   G   R   A   Q   L   R   R   S   L     303
AGC TGC CTC AAT CCC ATG CTC TTC CTG TAT TTT GGG AGG GCT CAA CTC CGC CGG TCA CTG     909

 P   A   A   C   H   W   A   L   R   E   S   Q   G   Q   D   E   S   V   D   S     323
CCA GCT GCC TGT CAC TGG GCC CTG AGG GAG TCC CAG GGC CAG GAC GAA AGT GTG GAC AGC     969

 K   K   S   T   S   H   D   L   V   S   E   M   E   V   *                       338
AAG AAA TCC ACC AGC CAT GAC CTG GTC TCG GAG ATG GAG GTG TAG                      1014   1080

GCTGGAGAGACATTGTGGGTGTGTATCTTCTTATCTCATTTCACAAGACTGGCTTCAGGCATAGCTGGATCCAGGAGCT

CAATGATGTCTTCATTTTATTCCTTCCTTCATTCAACAGATATCCATCATGCACTTGCTATGTGCAAGGCCTTTTTAGG
```

FIG 8

CACTAGAGATATAGCAGTGACCAAAACAG____AAATCCTGCCCTCAGGGAGCTGATATTCTTCTAGTGGAGGA CA

GACTATAAACAAAGATA

FIG 8 (CONT)

BNSDOCID: <WO 0149847A2 | >

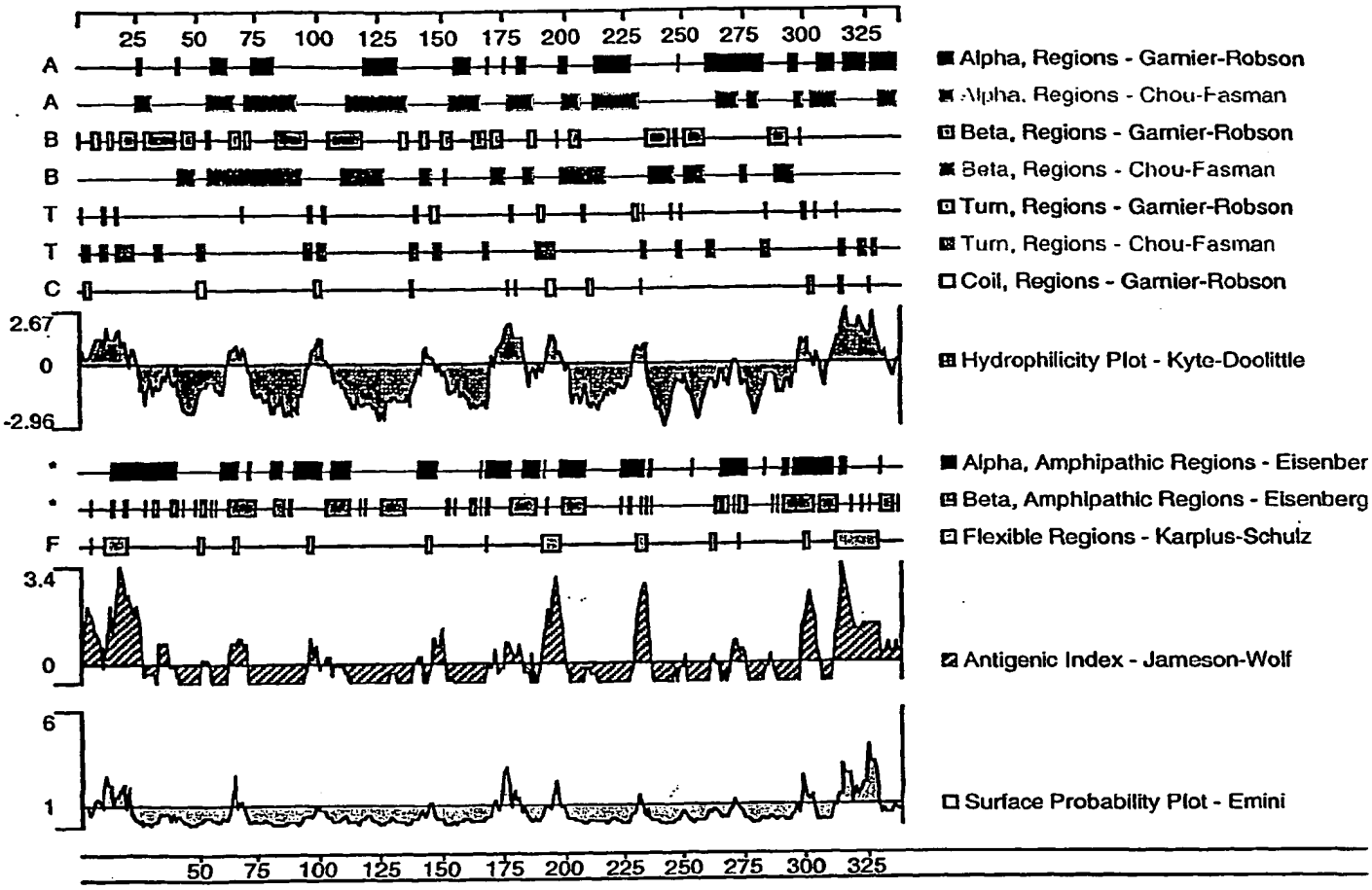

25 50 75 100 125 150 175 200 225 250 275 300 325
A
A
B
B
T
T
C
2.67
0
-2.96
*
*
F
3.4
0
6
1
50 75 100 125 150 175 200 225 250 275 300 325
Alpha, Regions - Garnier-Robson
Alpha, Regions - Chou-Fasman
Beta, Regions - Garnier-Robson
Beta, Regions - Chou-Fasman
Turn, Regions - Garnier-Robson
Turn, Regions - Chou-Fasman
Coil, Regions - Garnier-Robson
Hydrophilicity Plot - Kyte-Doolittle
Alpha, Amphipathic Regions - Eisenber
Beta, Amphipathic Regions - Eisenberg
Flexible Regions - Karplus-Schulz
Antigenic Index - Jameson-Wolf
Surface Probability Plot - Emini

FIG 9

Analysis of 38911 (337 aa)

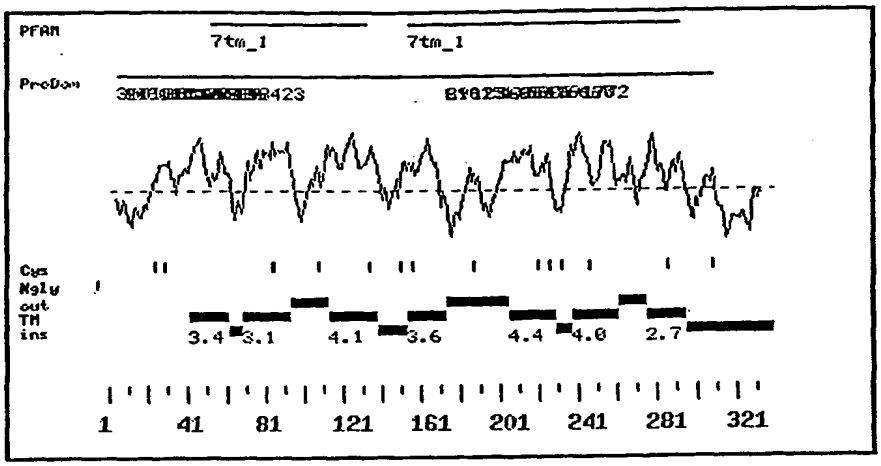


FIG 10

Transmembrane Segmer Predicted by MEMSAT

Start	End	Orient	Score
41	60	out-->ins	3.4
68	92	ins-->out	3.1
113	137	out-->ins	4.1
153	172	ins-->out	3.6
205	228	out-->ins	4.4
237	260	ins-->out	4.0
275	294	out-->ins	2.7

Prosite Pattern Matches for 38911

>PS00001|PDOC00001|ASN_GLYCOSYLATION N-glycosylation site.

Query: 3 NDSV 6

>PS00004|PDOC00004|CAMP_PHOSPHO_SITE cAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 324 KKST 327

>PS00005|PDOC00005|PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 17 SDR 19

Query: 323 SKK 325

>PS00006|PDOC00006|CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query: 194 SSTE 197

Query: 327 TSHD 330

Query: 333 SEME 336

>PS00008|PDOC00008|MYRISTYL N-myristoylation site.

Query: 26 GACLAI 31

Query: 49 GVPGNA 54

Query: 103 GAVGCR 108

Query: 150 GVQVAC 155

Query: 156 GAAWTL 161

Query: 191 GGSSST 196

Query: 253 GLVLTV 258

Query: 278 GLALAH 283

Query: 316 GQDESV 321

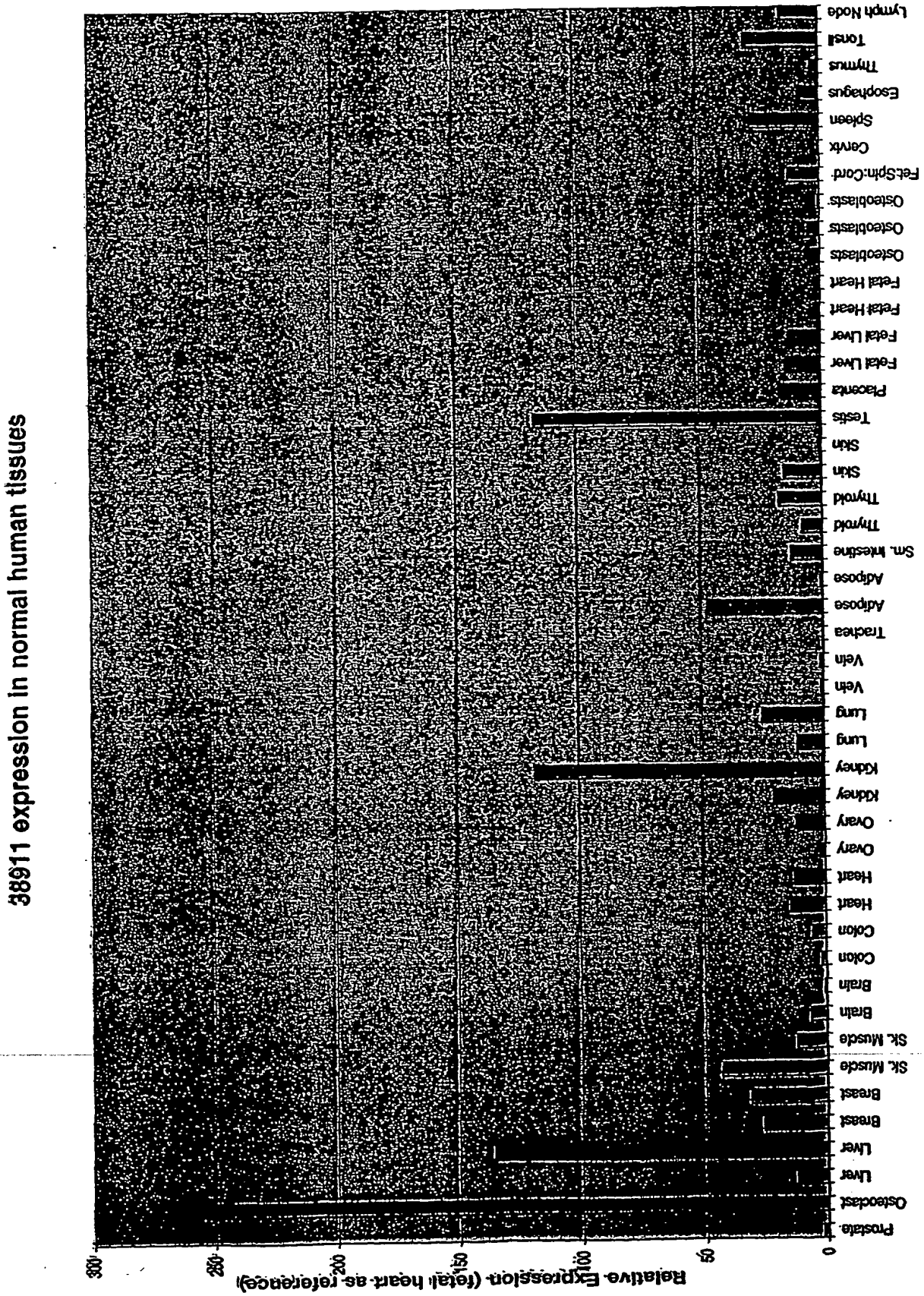
>PS00013|PDOC00013|PROKAR_LIPOPROTEIN Prokaryotic membrane lipoprotein lipid attachment site.
Non-eukaryotic pattern
RU Additional rules:
RU (1) The cysteine must be between positions 15 and 35 of the sequence in
RU consideration.
RU (2) There must be at least one charged residue (Lys or Arg) in the first
RU seven residues of the sequence.

Query: 97 GGHWPYGAVGC 107

Query: 209 FLGPLVAVASC 219

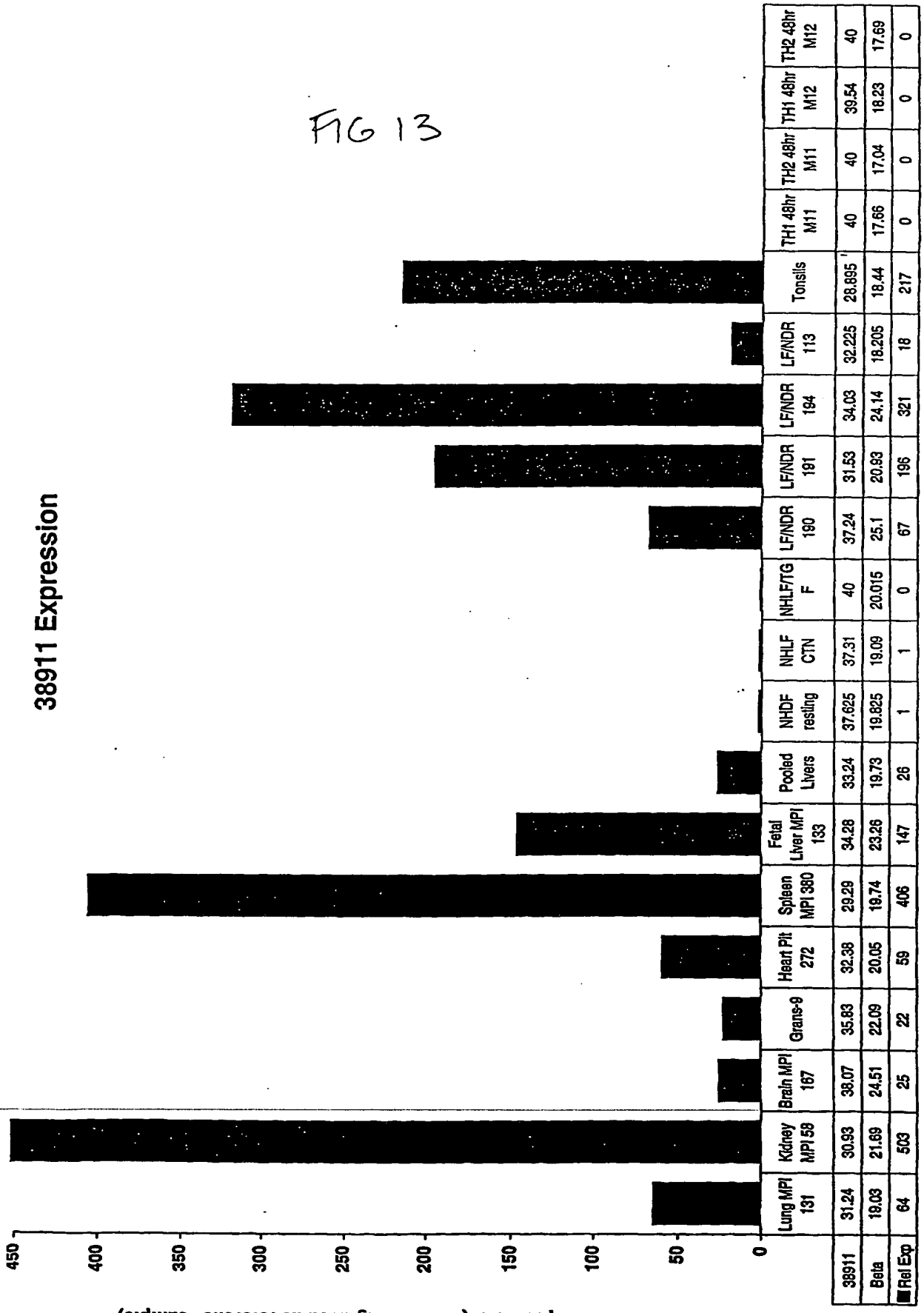
FIG 11

FIG 12



BNSDOCID: <WO___0149847A2_I_>

FIG 13



	Lung MPI 131	Kidney MPI 58	Brain MPI 167	Grans-9	Heart Plt 272	Spleen MPI 380	Fetal Liver MPI 133	Pooled Livers	NHDF resting	NHLF CTN	NHLF/TG F	LF/NDR 190	LF/NDR 191	LF/NDR 194	LF/NDR 113	Tonsils	TH1 48hr M11	TH2 48hr M11	TH1 48hr M12	TH2 48hr M12
38911	31.24	30.93	38.07	35.83	32.38	29.29	34.28	33.24	37.625	37.31	40	37.24	31.53	34.03	32.225	28.895	40	40	39.54	40
Beta	19.03	21.69	24.51	22.09	20.05	19.74	23.26	19.73	19.825	19.09	20.015	25.1	20.93	24.14	18.205	18.44	17.66	17.04	18.23	17.69
Rel Exp	64	503	25	22	59	406	147	26	1	1	0	67	196	321	18	217	0	0	0	0

Input file fbh26904.seq: Output File 26904.trans
Sequence length 1743

```
                10        20        30        40        50        60        70
GGCAGTGCACGCTCAGACGCCCCGCTCCTCCCGCCAGCGCGCGGCCTCGCTCCTCCTAGAGGACGCTCTCTGCGCGGGC
                80        90       100       110       120       130       140       150
CCTCGGAGGAGGCGGCGGCGGGGCGAGCTGCAGCGCCGGGACAGGAGGTTTGTCCCCGCCCGCGCGCCGTACCGCGGCG

                160
                M   G   E   T   M   S   K   R   V   R   L   H   L   G   G   E   A   E   M    19
GAG ATG GGC GAG ACC ATG TCA AAA CGC GTC CGG CTC CAC CTG GGA GGG GAG GCA GAA ATG    57  218
                1
                162
    E   E   R   A   F   V   N   P   F   P   D   Y   E   A   A   A   G   A   L   L    39
GAG GAA CGG GCG TTC GTC AAC CCC TTC CCG GAC TAC GAG GCC GCC GCC GGG GCG CTG CTC    117  278
    A   S   G   A   A   E   E   T   G   C   V   R   P   P   A   T   T   D   E   P    59
GCC TCC GGA GCG GCC GAA GAG ACA GGC TGT GTT CGT CCC CCG GCG ACC ACG GAT GAG CCC    177  338
    G   L   P   F   H   Q   D   G   K   I   I   H   N   F   I   R   R   I   Q   T    79
GGC CTC CCT TTT CAT CAG GAC GGG AAG ATC ATT CAT AAT TTC ATA AGA CGG ATC CAG ACC    237  398
    K   I   K   D   L   L   Q   Q   M   E   E   G   L   K   T   A   D   P   H   D    99
AAA ATT AAA GAT CTT CTG CAG CAA ATG GAA GAA GGG CTG AAG ACA GCT GAT CCC CAT GAC    297  458
    C   S   A   Y   T   G   W   T   G   I   A   L   L   Y   L   Q   L   Y   R   V   119
TGC TCT GCT TAT ACT GGC TGG ACA GGC ATA GCC CTT TTG TAC CTG CAG TTG TAC CGG GTC    357  518
    T   C   D   Q   T   Y   L   L   R   S   L   D   Y   V   K   R   T   L   R   N   139
ACA TGT GAC CAA ACC TAC CTG CTC CGA TCC CTG GAT TAC GTA AAA AGA ACA CTT CGG AAT    417  578
    L   N   G   R   R   V   T   F   L   C   G   D   A   G   P   L   A   V   G   A   159
CTG AAT GGC CGC AGG GTC ACC TTC CTC TGT GGG GAT GCT GGC CCC CTG GCT GTT GGA GCT    477  638
    V   I   Y   H   K   L   R   S   D   C   E   S   Q   E   C   V   T   K   L   L   179
GTG ATT TAT CAC AAA CTC AGA AGT GAC TGT GAG TCC CAG GAA TGT GTC ACA AAA CTT TTG    537  698
    Q   L   Q   R   S   V   V   C   Q   E   S   D   L   P   D   E   L   L   Y   G   199
CAG CTC CAG AGA TCG GTT GTC TGC CAA GAA TCA GAC CTT CCT GAT GAG CTG CTT TAT GGA    597  758
    R   A   G   Y   L   Y   A   L   L   Y   L   N   T   E   I   G   P   G   T   V   219
CGG GCA GGT TAT CTG TAT GCC TTA CTG TAC CTG AAC ACA GAG ATA GGT CCA GGC ACC GTG    657  818
    C   E   S   A   I   K   E   V   V   N   A   I   I   E   S   G   K   T   L   S   239
TGT GAG TCA GCT ATT AAA GAG GTA GTC AAT GCT ATT ATT GAA TCG GGT AAG ACT TTG TCA    717  877
    R   E   E   R   K   T   E   R   C   P   L   L   Y   Q   W   H   R   K   Q   Y   259
AGG GAA GAA AGA AAA ACG GAG CGC TGC CCG CTG TTG TAC CAG TGG CAC CGG AAG CAG TAC    777  937
    V   G   A   A   H   G   M   A   G   I   Y   Y   M   L   M   Q   P   A   A   K   279
GTT GGA GCA GCC CAT GGC ATG GCT GGA ATT TAC TAT ATG TTA ATG CAG CCG GCA GCA AAA    837  998
    V   D   Q   E   T   L   T   E   M   V   K   P   S   I   D   Y   V   R   H   K   299
GTG GAC CAA GAA ACC TTG ACA GAA ATG GTG AAA CCC AGT ATT GAT TAT GTG CGC CAC AAA    897  1058
    K   F   R   S   G   N   Y   P   S   S   L   S   N   E   T   D   R   L   V   H   319
AAA TTC CGA TCT GGG AAT TAC CCA TCA TCA TTA AGC AAT GAA ACA GAC CGG CTG GTG CAC    957  1118
    W   C   H   G   A   P   G   V   I   H   M   L   M   Q   A   Y   K   V   F   K   339
```

FIG 14

```
TGG TGC CAC GGC GCC CCG     GTC ATC CAC ATG CTC ATG CAG GCG T    AAG GTC TTT AAG  1017  11
 E   E   K   Y   L   K   E   A   M   E   C   S   D   V   I   W   Q   R   G   L    359
GAG GAG AAG TAC TTG AAA GAG GCC ATG GAG TGT AGC GAT GTG ATT TGG CAG CGA GGT TTG  1077  12
 L   R   K   G   Y   G   I   C   H   G   T   A   G   H   G   Y   S   F   L   S    379
CTG CGG AAG GGC TAC GGG ATA TGC CAT GGG ACT GCT GGC CAC GGC TAT TCC TTC CTG TCC  1137  129
 L   Y   R   L   T   Q   D   K   K   Y   L   Y   R   A   C   K   F   A   E   W    399
CTT TAC CGT CTC ACG CAG GAT AAG AAG TAC CTC TAC CGA GCT TGC AAG TTT GCA GAG TGG  1197  135
 C   L   D   Y   G   A   H   G   C   R   I   P   D   R   P   Y   S   L   F   E    419
TGT CTA GAT TAC GGA GCA CAC GGG TGC CGC ATT CCT GAC AGA CCC TAT TCG CTC TTT GAA  1257  141
 G   M   A   G   A   I   H   F   L   S   D   V   L   G   P   E   T   S   R   F    439
GGC ATG GCT GGC GCT ATT CAC TTT CTC TCT GAT GTC CTG GGA CCA GAG ACA TCA CGG TTT  1317  147
 P   A   F   E   L   D   S   S   K   R   D   *                                    451
CCA GCA TTT GAA CTT GAC TCT TCG AAG AGG GAT TAA                                  1353  151

AAGGTGCAAAAAGACAACTAAAATACCCATTTGGACCAAAAGCCGCCAGATTGCTTAGTGCCTGACACAGAAACAACTG

GGAATCCTGAAAGAGAAGCAGACACCGTCACAGGCCCCTCTGGTTAGACTAGCATGAGTGACCGAAGCCATCCATCAAC

ATTTTCTAACAGCACCCTCATCAATATAAAATATGACTTCTTCACATACAAAAAAAAAAAAAAAAGGGCGG
```

FIG 14 (CONT)

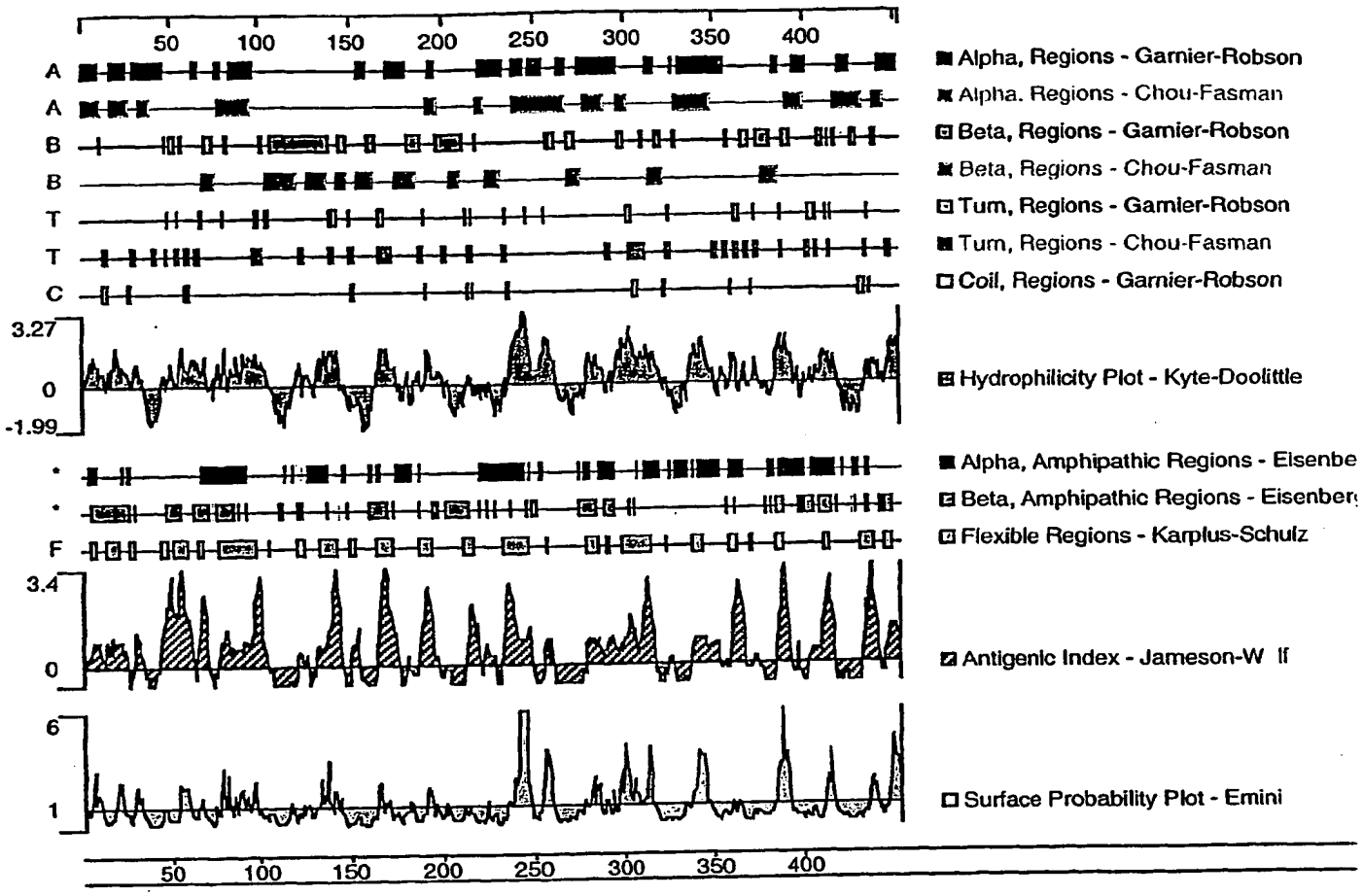

50
100
150
200
250
300
350
400
A
A
B
B
T
T
C
3.27
0
-1.99
*
*
F
3.4
0
6
1
Alpha, Regions - Garnier-Robson
Alpha, Regions - Chou-Fasman
Beta, Regions - Garnier-Robson
Beta, Regions - Chou-Fasman
Turn, Regions - Garnier-Robson
Turn, Regions - Chou-Fasman
Coil, Regions - Garnier-Robson
Hydrophilicity Plot - Kyte-Doolittle
Alpha, Amphipathic Regions - Eisenbe
Beta, Amphipathic Regions - Eisenber
Flexible Regions - Karplus-Schulz
Antigenic Index - Jameson-W lf
Surface Probability Plot - Emini
50
100
150
200
250
300
350
400

FIG 15

Analysis of 26904 (450 aa)

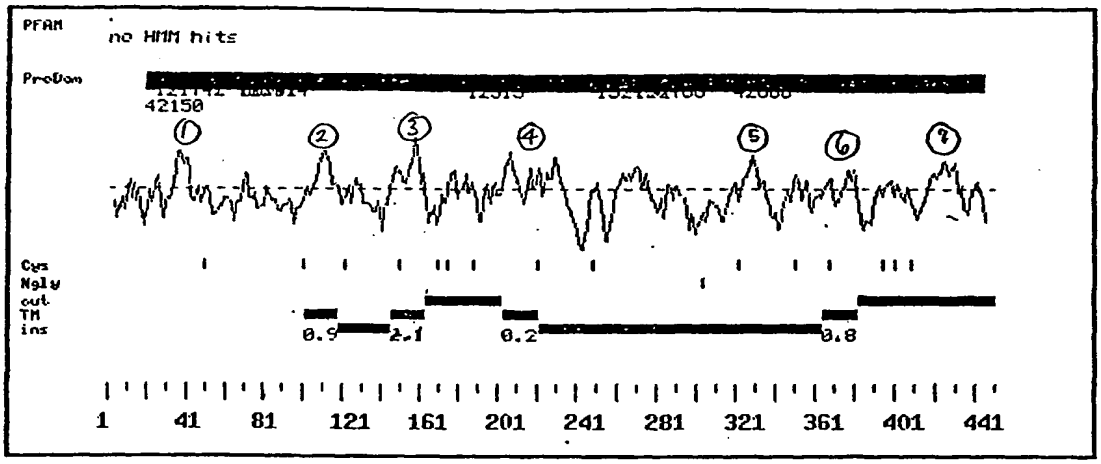


FIG 16

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	
101	117	out-->ins	0.9
145	162	ins-->out	2.1
203	220	out-->ins	0.2
364	381	ins-->out	0.8

Prosite Pattern Matches for 26904

>PS00001|PDOC00001|ASN_GLYCOSYLATION N-glycosylation site.

Query: 312 NETD 315

>PS00004|PDOC00004|CAMP_PHOSPHO_SITE cAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 143 RRVT 146

>PS00005|PDOC00005|PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 6 SKR 8
Query: 136 TLR 138
Query: 234 SGK 236
Query: 245 TER 247
Query: 314 TDR 316
Query: 436 TSR 438
Query: 446 SSK 448

>PS00006|PDOC00006|CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query: 55 TTDE 58
Query: 167 SDCE 170
Query: 218 TVCE 221
Query: 239 SREE 242
Query: 284 TLTE 287
Query: 416 SLFE 419
Query: 447 SKRD 450

>PS00007|PDOC00007|TYR_PHOSPHO_SITE Tyrosine kinase phosphorylation site.

Query: 118 RVTCDQTY 125
Query: 336 KVFKEEKY 343
Query: 382 RLTQDKKY 389
Query: 409 RIPDRPY 415

>PS00008|PDOC00008|MYRISTYL N-myristoylation site.

Query: 36 GALLAS 41
Query: 91 GLKTAD 96
Query: 261 GAAHGM 266
Query: 304 GNYPSS 309
Query: 365 GICHGT 370
Query: 404 GAHGCR 409
Query: 420 GMAGAI 425

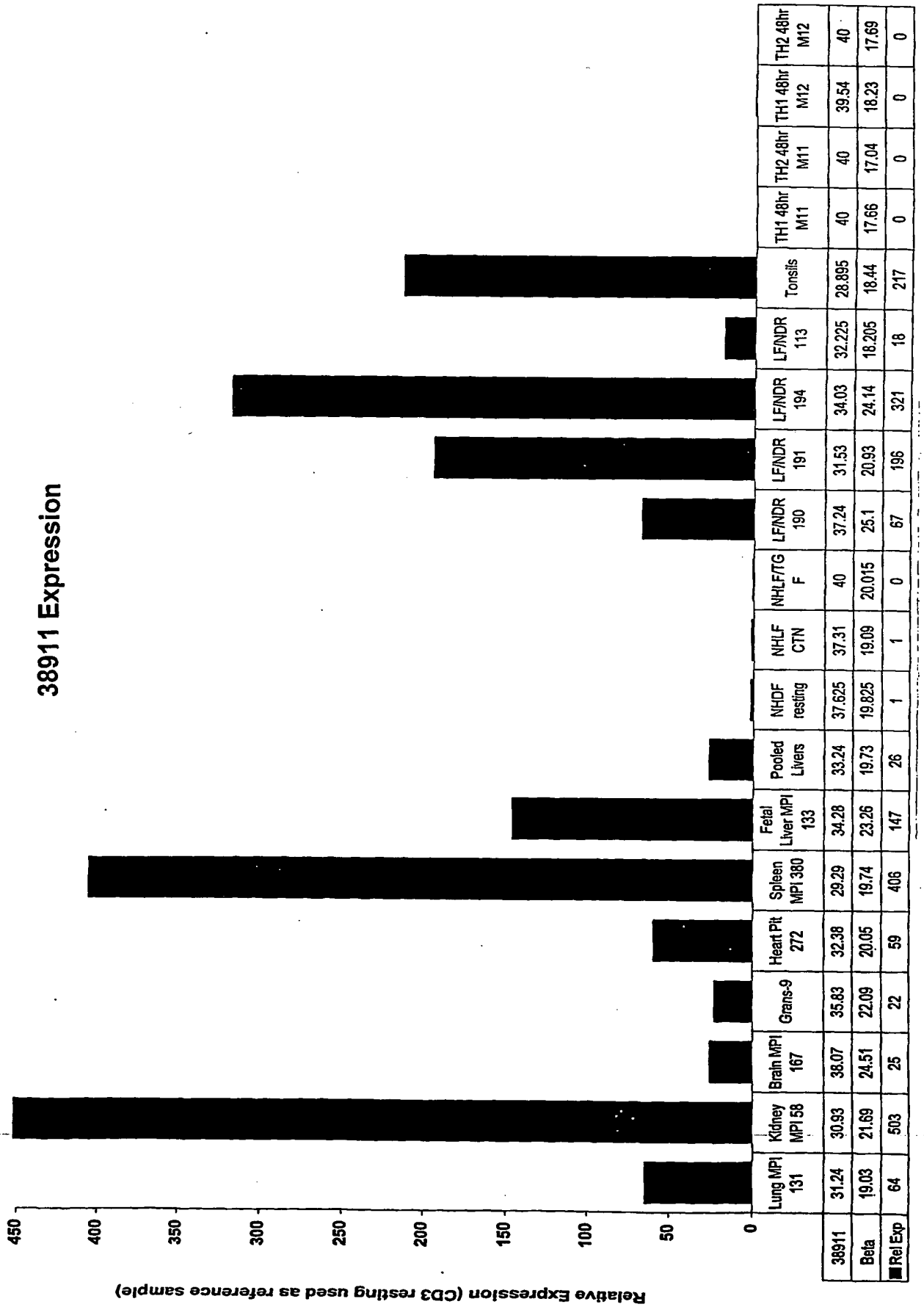
>PS00009|PDOC00009|AMIDATION Amidation site.

Query: 141 NGRR 144

>PS00017|PDOC00017|ATP_GTP_A ATP/GTP-binding site motif A (P-loop).

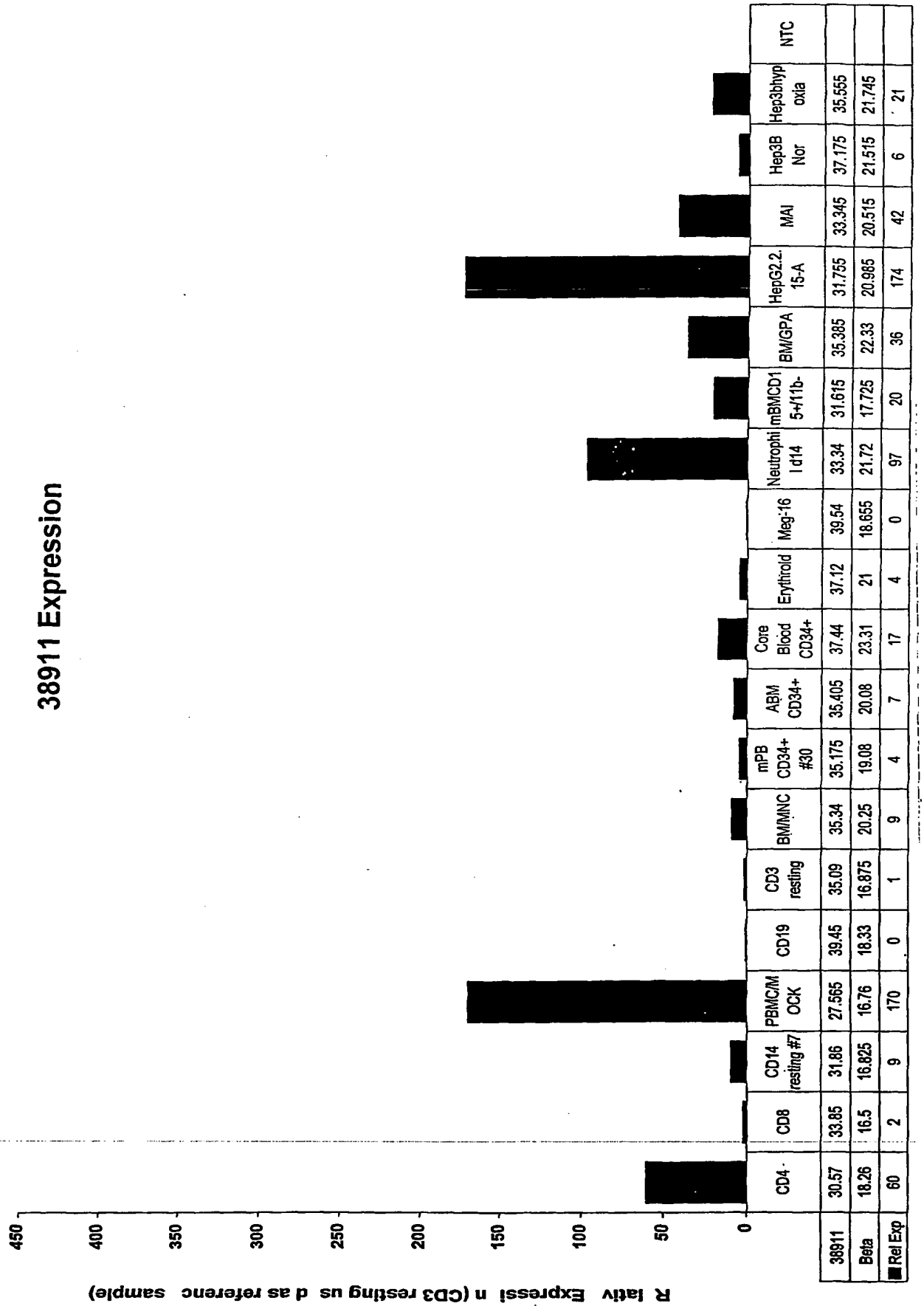
Query: 230 AIIESGKT 237

FIG 17



	Lung MPI 131	Kidney MPI 58	Brain MPI 167	Grans-9	Heart Plt 272	Spleen MPI 380	Fetal Liver MPI 133	Pooled Livers	NHDF resting	NHLF CTN	NHLF/TG F	LF/NDR 190	LF/NDR 191	LF/NDR 194	LF/NDR 113	Tonsils	TH1 48hr M11	TH2 48hr M11	TH1 48hr M12	TH2 48hr M12
38911	31.24	30.93	38.07	35.83	32.38	29.29	34.28	33.24	37.625	37.31	40	37.24	31.53	34.03	32.225	28.895	40	40	39.54	40
Beta	19.03	21.69	24.51	22.09	20.05	19.74	23.26	19.73	19.825	19.09	20.015	25.1	20.93	24.14	18.205	18.44	17.66	17.04	18.23	17.69
Rel Exp	64	503	25	22	59	406	147	26	1	1	0	67	196	321	18	217	0	0	0	0

FIGURE 18



	CD4	CD8	CD14 resting #7	PBMC/MOCK	CD19	CD3 resting	BM/MNC	mPB CD34+ #30	ABM CD34+	Core Blood CD34+	Erythroid	Meg-16	Neutrophil d14	mBMCD15+/11b-	BM/GPA	HepG2.2.15-A	MAI	Hep3B Nor	Hep3bhypoxia	NTC
38911	30.57	33.85	31.86	27.565	39.45	35.09	35.34	35.175	35.405	37.44	37.12	39.54	33.34	31.615	35.385	31.755	33.345	37.175	35.555	
Beta	18.26	16.5	16.825	16.76	18.33	16.875	20.25	19.08	20.08	23.31	21	18.655	21.72	17.725	22.33	20.985	20.515	21.515	21.745	
Rel Exp	60	2	9	170	0	1	9	4	7	17	4	0	97	20	36	174	42	6	21	

FIGURE 19

SEQUENCE LISTING

<110> Glucksmann, Maria Alexandra
White, David

<120> 26904, 38911, and 39404, Novel
Seven-Transmembrane Proteins/G-Protein Coupled Receptors

<130> 35800/207180

<160> 34

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 337
<212> PRT
<213> Homo sapiens

<400> 1

```
Met Asn Glu Pro Leu Asp Tyr Leu Ala Asn Ala Ser Asp Phe Pro Asp
 1               5                  10                  15
Tyr Ala Ala Ala Phe Gly Asn Cys Thr Asp Glu Asn Ile Pro Leu Lys
            20                  25                  30
Met His Tyr Leu Pro Val Ile Tyr Gly Ile Ile Phe Leu Val Gly Phe
        35                  40                  45
Pro Gly Asn Ala Val Val Ile Ser Thr Tyr Ile Phe Lys Met Arg Pro
    50                  55                  60
Trp Lys Ser Ser Thr Ile Ile Met Leu Asn Leu Ala Cys Thr Asp Leu
65                  70                  75                  80
Leu Tyr Leu Thr Ser Leu Pro Phe Leu Ile His Tyr Tyr Ala Ser Gly
                85                  90                  95
Glu Asn Trp Ile Phe Gly Asp Phe Met Cys Lys Phe Ile Arg Phe Ser
            100                 105                 110
Phe His Phe Asn Leu Tyr Ser Ser Ile Leu Phe Leu Thr Cys Phe Ser
        115                 120                 125
Ile Phe Arg Tyr Cys Val Ile Ile His Pro Met Ser Cys Phe Ser Ile
    130                 135                 140
His Lys Thr Arg Cys Ala Val Val Ala Cys Ala Val Val Trp Ile Ile
145                 150                 155                 160
Ser Leu Val Ala Val Ile Pro Met Thr Phe Leu Ile Thr Ser Thr Asn
                165                 170                 175
Arg Thr Asn Arg Ser Ala Cys Leu Asp Leu Thr Ser Ser Asp Glu Leu
            180                 185                 190
Asn Thr Ile Lys Trp Tyr Asn Leu Ile Leu Thr Ala Thr Thr Phe Cys
        195                 200                 205
Leu Pro Leu Val Ile Val Thr Leu Cys Tyr Thr Thr Ile Ile His Thr
    210                 215                 220
Leu Thr His Gly Leu Gln Thr Asp Ser Cys Leu Lys Gln Lys Ala Arg
225                 230                 235                 240
Arg Leu Thr Ile Leu Leu Leu Leu Ala Phe Tyr Val Cys Phe Leu Pro
                245                 250                 255
Phe His Ile Leu Arg Val Ile Arg Ile Glu Ser Arg Leu Leu Ser Ile
            260                 265                 270
Ser Cys Ser Ile Glu Asn Gln Ile His Glu Ala Tyr Ile Val Ser Gly
        275                 280                 285
Pro Leu Ala Ala Leu Asn Thr Phe Gly Asn Leu Leu Leu Tyr Val Val
    290                 295                 300
Val Ser Asp Asn Phe Gln Gln Ala Val Cys Ser Thr Val Arg Cys Lys
305                 310                 315                 320
Val Ser Gly Asn Leu Glu Gln Ala Lys Lys Ile Ser Tyr Ser Asn Asn
                325                 330                 335
Pro
```

<210> 2
<211> 1729
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (294)...(1307)

<400> 2
```
cctttttttt ttttttttaa cttttatatt tttattagat gcatttagta acttgcctca      60
tagtcatttt cttggaaatt caatttcttc tccacagggt ctcttttgag attaaagaga     120
gagaagtggc aaatttagga tgttagaata attttcattt aaaagtagat ccttgttttt     180
attaccctat cattaatgtt ttctgttttc ctttatcagc gagttactgc tcatttgatt     240
catattgcca aactgaactc tcttgttttc ttgcaagatg aaaggagaca acc atg         296
                                                              Met
                                                               1

aat gag cca cta gac tat tta gca aat gct tct gat ttc ccc gat tat      344
Asn Glu Pro Leu Asp Tyr Leu Ala Asn Ala Ser Asp Phe Pro Asp Tyr
            5                  10                  15

gca gct gct ttt gga aat tgc act gat gaa aac atc cca ctc aag atg      392
Ala Ala Ala Phe Gly Asn Cys Thr Asp Glu Asn Ile Pro Leu Lys Met
        20                  25                  30

cac tac ctc cct gtt att tat ggc att atc ttc ctc gtg gga ttt cca      440
His Tyr Leu Pro Val Ile Tyr Gly Ile Ile Phe Leu Val Gly Phe Pro
    35                  40                  45

ggc aat gca gta gtg ata tcc act tac att ttc aaa atg aga cct tgg      488
Gly Asn Ala Val Val Ile Ser Thr Tyr Ile Phe Lys Met Arg Pro Trp
50                  55                  60                  65

aag agc agc acc atc att atg ctg aac ctg gcc tgc aca gat ctg ctg      536
Lys Ser Ser Thr Ile Ile Met Leu Asn Leu Ala Cys Thr Asp Leu Leu
                70                  75                  80

tat ctg acc agc ctc ccc ttc ctg att cac tac tat gcc agt ggc gaa      584
Tyr Leu Thr Ser Leu Pro Phe Leu Ile His Tyr Tyr Ala Ser Gly Glu
            85                  90                  95

aac tgg atc ttt gga gat ttc atg tgt aag ttt atc cgc ttc agc ttc      632
Asn Trp Ile Phe Gly Asp Phe Met Cys Lys Phe Ile Arg Phe Ser Phe
        100                 105                 110

cat ttc aac ctg tat agc agc atc ctc ttc ctc acc tgt ttc agc atc      680
His Phe Asn Leu Tyr Ser Ser Ile Leu Phe Leu Thr Cys Phe Ser Ile
    115                 120                 125

ttc cgc tac tgt gtg atc att cac cca atg agc tgc ttt tcc att cac      728
Phe Arg Tyr Cys Val Ile Ile His Pro Met Ser Cys Phe Ser Ile His
130                 135                 140                 145

aaa act cga tgt gca gtt gta gcc tgt gct gtg gtg tgg atc att tca      776
Lys Thr Arg Cys Ala Val Val Ala Cys Ala Val Val Trp Ile Ile Ser
                150                 155                 160

ctg gta gct gtc att ccg atg acc ttc ttg atc aca tca acc aac agg      824
Leu Val Ala Val Ile Pro Met Thr Phe Leu Ile Thr Ser Thr Asn Arg
            165                 170                 175

acc aac aga tca gcc tgt ctc gac ctc acc agt tcg gat gaa ctc aat      872
Thr Asn Arg Ser Ala Cys Leu Asp Leu Thr Ser Ser Asp Glu Leu Asn
        180                 185                 190

act att aag tgg tac aac ctg att ttg act gca act act ttc tgc ctc      920
```


BNSDOCID <WO___0149847A2_I_>

```
Thr Ile Lys Trp Tyr Asn Leu Ile Leu Thr Ala Thr Thr Phe Cys Leu
    195                 200                 205

ccc ttg gtg ata gtg aca ctt tgc tat acc acg att atc cac act ctg      968
Pro Leu Val Ile Val Thr Leu Cys Tyr Thr Thr Ile Ile His Thr Leu
210                 215                 220                 225

acc cat gga ctg caa act gac agc tgc ctt aag cag aaa gca cga agg     1016
Thr His Gly Leu Gln Thr Asp Ser Cys Leu Lys Gln Lys Ala Arg Arg
                230                 235                 240

cta acc att ctg cta ctc ctt gca ttt tac gta tgt ttt tta ccc ttc     1064
Leu Thr Ile Leu Leu Leu Leu Ala Phe Tyr Val Cys Phe Leu Pro Phe
            245                 250                 255

cat atc ttg agg gtc att cgg atc gaa tct cgc ctg ctt tca atc agt     1112
His Ile Leu Arg Val Ile Arg Ile Glu Ser Arg Leu Leu Ser Ile Ser
        260                 265                 270

tgt tcc att gag aat cag atc cat gaa gct tac atc gtt tct gga cca     1160
Cys Ser Ile Glu Asn Gln Ile His Glu Ala Tyr Ile Val Ser Gly Pro
    275                 280                 285

tta gct gct ctg aac acc ttt ggt aac ctg tta cta tat gtg gtg gtc     1208
Leu Ala Ala Leu Asn Thr Phe Gly Asn Leu Leu Leu Tyr Val Val Val
290                 295                 300                 305

agc gac aac ttt cag cag gct gtc tgc tca aca gtg aga tgc aaa gta     1256
Ser Asp Asn Phe Gln Gln Ala Val Cys Ser Thr Val Arg Cys Lys Val
                310                 315                 320

agc ggg aac ctt gag caa gca aag aaa att agt tac tca aac aac cct     1304
Ser Gly Asn Leu Glu Gln Ala Lys Lys Ile Ser Tyr Ser Asn Asn Pro
            325                 330                 335

tga aatatttcat ttacttaacc aaaaacaaat acttgctgat actttaccta          1357
 *


gcatcctaag atgttcagga tgtctccctc aatggaactc ctggtaaata ctgtgtattc  1417
aagtaatcat gtgccaaagc cagggcagag cttctagttc tttgcaatcc ctttattgag  1477
ctcctccact ggggagatat aagaatggga tgcatgtata tcagcaaagt attcagacat  1537
agtattacaa gctattggaa ctcagaggca tcttagagaa catctgttcc caccaactta  1597
ctatatatac acggaaacca atttcttacc cttgccctag attgctcagt aaatttgtgc  1657
caagatagga gaaaaccaat cttttcactc atcatttcat gcttctctgc actctgggcc  1717
tatttgtatt ga                                                      1729

<210> 3
<211> 337
<212> PRT
<213> Homo sapeins

<400> 3
Met Gly Asn Asp Ser Val Ser Tyr Glu Tyr Gly Asp Tyr Ser Asp Leu
 1               5                  10                  15
Ser Asp Arg Pro Val Asp Cys Leu Asp Gly Ala Cys Leu Ala Ile Asp
            20                  25                  30
Pro Leu Arg Val Ala Pro Leu Pro Leu Tyr Ala Ala Ile Phe Leu Val
        35                  40                  45
Gly Val Pro Gly Asn Ala Met Val Ala Trp Val Ala Gly Lys Val Ala
    50                  55                  60
Arg Arg Arg Val Gly Ala Thr Trp Leu Leu His Leu Ala Val Ala Asp
65                  70                  75                  80
Leu Leu Cys Cys Leu Ser Leu Pro Ile Leu Ala Val Pro Ile Ala Arg
                85                  90                  95
Gly Gly His Trp Pro Tyr Gly Ala Val Gly Cys Arg Ala Leu Pro Ser
            100                 105                 110
```

```
Ile Ile Leu Leu Thr Met Tyr Ala Ser Val Leu Leu Leu Ala Ala Leu
        115                 120                 125
Ser Ala Asp Leu Cys Phe Leu Ala Leu Gly Pro Ala Trp Trp Ser Thr
    130                 135                 140
Val Gln Arg Ala Cys Gly Val Gln Val Ala Cys Gly Ala Ala Trp Thr
145                 150                 155                 160
Leu Ala Leu Leu Leu Thr Val Pro Ser Ala Ile Tyr Arg Arg Leu His
                165                 170                 175
Gln Glu His Phe Pro Ala Arg Leu Gln Cys Val Val Asp Tyr Gly Gly
            180                 185                 190
Ser Ser Ser Thr Glu Asn Ala Val Thr Ala Ile Arg Phe Leu Phe Gly
        195                 200                 205
Phe Leu Gly Pro Leu Val Ala Val Ala Ser Cys His Ser Ala Leu Leu
    210                 215                 220
Cys Trp Ala Ala Arg Arg Cys Arg Pro Leu Gly Thr Ala Ile Val Val
225                 230                 235                 240
Gly Phe Phe Val Cys Trp Ala Pro Tyr His Leu Leu Gly Leu Val Leu
                245                 250                 255
Thr Val Ala Ala Pro Asn Ser Ala Leu Leu Ala Arg Ala Leu Arg Ala
            260                 265                 270
Glu Pro Leu Ile Val Gly Leu Ala Leu Ala His Ser Cys Leu Asn Pro
        275                 280                 285
Met Leu Phe Leu Tyr Phe Gly Arg Ala Gln Leu Arg Arg Ser Leu Pro
    290                 295                 300
Ala Ala Cys His Trp Ala Leu Arg Glu Ser Gln Gly Gln Asp Glu Ser
305                 310                 315                 320
Val Asp Ser Lys Lys Ser Thr Ser His Asp Leu Val Ser Glu Met Glu
                325                 330                 335
Val
```

<210> 4
<211> 1334
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (67)...(1080)

<400> 4

```
gtccgacgtg ctggacaaat cttaactcct caaggactcc caaaaccaga gacaccagga      60
gcctga atg ggg aac gat tct gtc agc tac gag tat ggg gat tac agc          108
       Met Gly Asn Asp Ser Val Ser Tyr Glu Tyr Gly Asp Tyr Ser
        1               5                   10

gac ctc tcg gac cgc cct gtg gac tgc ctg gat ggc gcc tgc ctg gcc         156
Asp Leu Ser Asp Arg Pro Val Asp Cys Leu Asp Gly Ala Cys Leu Ala
 15                  20                  25                  30

atc gac ccg ctg cgc gtg gcc ccg ctc cca ctg tat gcc gcc atc ttc         204
Ile Asp Pro Leu Arg Val Ala Pro Leu Pro Leu Tyr Ala Ala Ile Phe
                35                  40                  45

ctg gtg ggg gtg ccg ggc aat gcc atg gtg gcc tgg gtg gct ggg aag         252
Leu Val Gly Val Pro Gly Asn Ala Met Val Ala Trp Val Ala Gly Lys
            50                  55                  60

gtg gcc cgc cgg agg gtg ggt gcc acc tgg ttg ctc cac ctg gcc gtg         300
Val Ala Arg Arg Arg Val Gly Ala Thr Trp Leu Leu His Leu Ala Val
        65                  70                  75

gcg gat ttg ctg tgc tgt ttg tct ctg ccc atc ctg gca gtg ccc att         348
Ala Asp Leu Leu Cys Cys Leu Ser Leu Pro Ile Leu Ala Val Pro Ile
    80                  85                  90

gcc cgt gga ggc cac tgg ccg tat ggt gca gtg ggc tgt cgg gcg ctg         396
```

```
Ala Arg Gly Gly His Trp Pro Tyr Gly Ala Val Gly Cys Arg Ala Leu
 95                  100                 105                 110

ccc tcc atc atc ctg ctg acc atg tat gcc agc gtc ctg ctc ctg gca      444
Pro Ser Ile Ile Leu Leu Thr Met Tyr Ala Ser Val Leu Leu Leu Ala
                115                 120                 125

gct ctc agt gcc gac ctc tgc ttc ctg gct ctc ggg cct gcc tgg tgg      492
Ala Leu Ser Ala Asp Leu Cys Phe Leu Ala Leu Gly Pro Ala Trp Trp
            130                 135                 140

tct acg gtt cag cgg gcg tgc ggg gtg cag gtg gcc tgt ggg gca gcc      540
Ser Thr Val Gln Arg Ala Cys Gly Val Gln Val Ala Cys Gly Ala Ala
        145                 150                 155

tgg aca ctg gcc ttg ctg ctc acc gtg ccc tcc gcc atc tac cgc cgg      588
Trp Thr Leu Ala Leu Leu Leu Thr Val Pro Ser Ala Ile Tyr Arg Arg
    160                 165                 170

ctg cac cag gag cac ttc cca gcc cgg ctg cag tgt gtg gtg gac tac      636
Leu His Gln Glu His Phe Pro Ala Arg Leu Gln Cys Val Val Asp Tyr
175                 180                 185                 190

ggc ggc tcc tcc agc acc gag aat gcg gtg act gcc atc cgg ttt ctt      684
Gly Gly Ser Ser Ser Thr Glu Asn Ala Val Thr Ala Ile Arg Phe Leu
                195                 200                 205

ttt ggc ttc ctg ggg ccc ctg gtg gcc gtg gcc agc tgc cac agt gcc      732
Phe Gly Phe Leu Gly Pro Leu Val Ala Val Ala Ser Cys His Ser Ala
            210                 215                 220

ctc ctg tgc tgg gca gcc cga cgc tgc cgg ccg ctg ggc aca gcc att      780
Leu Leu Cys Trp Ala Ala Arg Arg Cys Arg Pro Leu Gly Thr Ala Ile
        225                 230                 235

gtg gtg ggg ttt ttt gtc tgc tgg gca ccc tac cac ctg ctg ggg ctg      828
Val Val Gly Phe Phe Val Cys Trp Ala Pro Tyr His Leu Leu Gly Leu
    240                 245                 250

gtg ctc act gtg gcg gcc ccg aac tcc gca ctc ctg gcc agg gcc ctg      876
Val Leu Thr Val Ala Ala Pro Asn Ser Ala Leu Leu Ala Arg Ala Leu
255                 260                 265                 270

cgg gct gaa ccc ctc atc gtg ggc ctt gcc ctc gct cac agc tgc ctc      924
Arg Ala Glu Pro Leu Ile Val Gly Leu Ala Leu Ala His Ser Cys Leu
                275                 280                 285

aat ccc atg ctc ttc ctg tat ttt ggg agg gct caa ctc cgc cgg tca      972
Asn Pro Met Leu Phe Leu Tyr Phe Gly Arg Ala Gln Leu Arg Arg Ser
            290                 295                 300

ctg cca gct gcc tgt cac tgg gcc ctg agg gag tcc cag ggc cag gac     1020
Leu Pro Ala Ala Cys His Trp Ala Leu Arg Glu Ser Gln Gly Gln Asp
        305                 310                 315

gaa agt gtg gac agc aag aaa tcc acc agc cat gac ctg gtc tcg gag     1068
Glu Ser Val Asp Ser Lys Lys Ser Thr Ser His Asp Leu Val Ser Glu
    320                 325                 330

atg gag gtg tag gctggagaga cattgtgggt gtgtatcttc ttatctcatt         1120
Met Glu Val  *
335

tcacaagact ggcttcaggc atagctggat ccaggagctc aatgatgtct tcattttatt   1180
ccttccttca ttcaacagat atccatcatg cacttgctat gtgcaaggcc tttttaggca   1240
ctagagatat agcagtgacc aaaacagaca caaatcctgc cctcagggag ctgatattct   1300
tctagtggag gaagacagac tataaacaaa gata                               1334
```

<210> 5
<211> 450
<212> PRT
<213> Homo sapiens

<400> 5

```
Met Gly Glu Thr Met Ser Lys Arg Val Arg Leu His Leu Gly Gly Glu
 1               5                  10                  15
Ala Glu Met Glu Glu Arg Ala Phe Val Asn Pro Phe Pro Asp Tyr Glu
            20                  25                  30
Ala Ala Ala Gly Ala Leu Leu Ala Ser Gly Ala Ala Glu Glu Thr Gly
        35                  40                  45
Cys Val Arg Pro Pro Ala Thr Thr Asp Glu Pro Gly Leu Pro Phe His
    50                  55                  60
Gln Asp Gly Lys Ile Ile His Asn Phe Ile Arg Arg Ile Gln Thr Lys
65                  70                  75                  80
Ile Lys Asp Leu Leu Gln Gln Met Glu Glu Gly Leu Lys Thr Ala Asp
                85                  90                  95
Pro His Asp Cys Ser Ala Tyr Thr Gly Trp Thr Gly Ile Ala Leu Leu
            100                 105                 110
Tyr Leu Gln Leu Tyr Arg Val Thr Cys Asp Gln Thr Tyr Leu Leu Arg
        115                 120                 125
Ser Leu Asp Tyr Val Lys Arg Thr Leu Arg Asn Leu Asn Gly Arg Arg
    130                 135                 140
Val Thr Phe Leu Cys Gly Asp Ala Gly Pro Leu Ala Val Gly Ala Val
145                 150                 155                 160
Ile Tyr His Lys Leu Arg Ser Asp Cys Glu Ser Gln Glu Cys Val Thr
                165                 170                 175
Lys Leu Leu Gln Leu Gln Arg Ser Val Val Cys Gln Glu Ser Asp Leu
            180                 185                 190
Pro Asp Glu Leu Leu Tyr Gly Arg Ala Gly Tyr Leu Tyr Ala Leu Leu
        195                 200                 205
Tyr Leu Asn Thr Glu Ile Gly Pro Gly Thr Val Cys Glu Ser Ala Ile
    210                 215                 220
Lys Glu Val Val Asn Ala Ile Ile Glu Ser Gly Lys Thr Leu Ser Arg
225                 230                 235                 240
Glu Glu Arg Lys Thr Glu Arg Cys Pro Leu Leu Tyr Gln Trp His Arg
                245                 250                 255
Lys Gln Tyr Val Gly Ala Ala His Gly Met Ala Gly Ile Tyr Tyr Met
            260                 265                 270
Leu Met Gln Pro Ala Ala Lys Val Asp Gln Glu Thr Leu Thr Glu Met
        275                 280                 285
Val Lys Pro Ser Ile Asp Tyr Val Arg His Lys Lys Phe Arg Ser Gly
    290                 295                 300
Asn Tyr Pro Ser Ser Leu Ser Asn Glu Thr Asp Arg Leu Val His Trp
305                 310                 315                 320
Cys His Gly Ala Pro Gly Val Ile His Met Leu Met Gln Ala Tyr Lys
                325                 330                 335
Val Phe Lys Glu Glu Lys Tyr Leu Lys Glu Ala Met Glu Cys Ser Asp
            340                 345                 350
Val Ile Trp Gln Arg Gly Leu Leu Arg Lys Gly Tyr Gly Ile Cys His
        355                 360                 365
Gly Thr Ala Gly His Gly Tyr Ser Phe Leu Ser Leu Tyr Arg Leu Thr
    370                 375                 380
Gln Asp Lys Lys Tyr Leu Tyr Arg Ala Cys Lys Phe Ala Glu Trp Cys
385                 390                 395                 400
Leu Asp Tyr Gly Ala His Gly Cys Arg Ile Pro Asp Arg Pro Tyr Ser
                405                 410                 415
Leu Phe Glu Gly Met Ala Gly Ala Ile His Phe Leu Ser Asp Val Leu
            420                 425                 430
Gly Pro Glu Thr Ser Arg Phe Pro Ala Phe Glu Leu Asp Ser Ser Lys
        435                 440                 445
Arg Asp
    450
```

<210> 6

RNSDOCID WO 0149847A2 | >

<211> 1743
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (162)...(1514)

<400> 6
```
ggcagtgcac gctcagacgc cccgctcctc ccgccagcgc gcggcctcgc tcctcctaga      60
ggacgctctc tgcgcgggcc ctcggaggag gcggcggcgg ggcgagctgc agcgccggga     120
caggaggttt gtccccgccc gcgcgccgta ccgcggcgga g atg ggc gag acc atg     176
                                                 Met Gly Glu Thr Met
                                                  1               5

tca aaa cgc gtc cgg ctc cac ctg gga ggg gag gca gaa atg gag gaa       224
Ser Lys Arg Val Arg Leu His Leu Gly Gly Glu Ala Glu Met Glu Glu
                 10                  15                  20

cgg gcg ttc gtc aac ccc ttc ccg gac tac gag gcc gcc gcc ggg gcg       272
Arg Ala Phe Val Asn Pro Phe Pro Asp Tyr Glu Ala Ala Ala Gly Ala
             25                  30                  35

ctg ctc gcc tcc gga gcg gcc gaa gag aca ggc tgt gtt cgt ccc ccg       320
Leu Leu Ala Ser Gly Ala Ala Glu Glu Thr Gly Cys Val Arg Pro Pro
         40                  45                  50

gcg acc acg gat gag ccc ggc ctc cct ttt cat cag gac ggg aag atc       368
Ala Thr Thr Asp Glu Pro Gly Leu Pro Phe His Gln Asp Gly Lys Ile
     55                  60                  65

att cat aat ttc ata aga cgg atc cag acc aaa att aaa gat ctt ctg       416
Ile His Asn Phe Ile Arg Arg Ile Gln Thr Lys Ile Lys Asp Leu Leu
 70                  75                  80                  85

cag caa atg gaa gaa ggg ctg aag aca gct gat ccc cat gac tgc tct       464
Gln Gln Met Glu Glu Gly Leu Lys Thr Ala Asp Pro His Asp Cys Ser
                 90                  95                 100

gct tat act ggc tgg aca ggc ata gcc ctt ttg tac ctg cag ttg tac       512
Ala Tyr Thr Gly Trp Thr Gly Ile Ala Leu Leu Tyr Leu Gln Leu Tyr
            105                 110                 115

cgg gtc aca tgt gac caa acc tac ctg ctc cga tcc ctg gat tac gta       560
Arg Val Thr Cys Asp Gln Thr Tyr Leu Leu Arg Ser Leu Asp Tyr Val
        120                 125                 130

aaa aga aca ctt cgg aat ctg aat ggc cgc agg gtc acc ttc ctc tgt       608
Lys Arg Thr Leu Arg Asn Leu Asn Gly Arg Arg Val Thr Phe Leu Cys
    135                 140                 145

ggg gat gct ggc ccc ctg gct gtt gga gct gtg att tat cac aaa ctc       656
Gly Asp Ala Gly Pro Leu Ala Val Gly Ala Val Ile Tyr His Lys Leu
150                 155                 160                 165

aga agt gac tgt gag tcc cag gaa tgt gtc aca aaa ctt ttg cag ctc       704
Arg Ser Asp Cys Glu Ser Gln Glu Cys Val Thr Lys Leu Leu Gln Leu
                170                 175                 180

cag aga tcg gtt gtc tgc caa gaa tca gac ctt cct gat gag ctg ctt       752
Gln Arg Ser Val Val Cys Gln Glu Ser Asp Leu Pro Asp Glu Leu Leu
            185                 190                 195

tat gga cgg gca ggt tat ctg tat gcc tta ctg tac ctg aac aca gag       800
Tyr Gly Arg Ala Gly Tyr Leu Tyr Ala Leu Leu Tyr Leu Asn Thr Glu
        200                 205                 210
```

```
ata ggt cca ggc acc gtg tgt gag tca gct att aaa gag gta gtc aat      848
Ile Gly Pro Gly Thr Val Cys Glu Ser Ala Ile Lys Glu Val Val Asn
    215                 220                 225

gct att att gaa tcg ggt aag act ttg tca agg gaa gaa aga aaa acg      896
Ala Ile Ile Glu Ser Gly Lys Thr Leu Ser Arg Glu Glu Arg Lys Thr
230                 235                 240                 245

gag cgc tgc ccg ctg ttg tac cag tgg cac cgg aag cag tac gtt gga      944
Glu Arg Cys Pro Leu Leu Tyr Gln Trp His Arg Lys Gln Tyr Val Gly
                250                 255                 260

gca gcc cat ggc atg gct gga att tac tat atg tta atg cag ccg gca      992
Ala Ala His Gly Met Ala Gly Ile Tyr Tyr Met Leu Met Gln Pro Ala
            265                 270                 275

gca aaa gtg gac caa gaa acc ttg aca gaa atg gtg aaa ccc agt att     1040
Ala Lys Val Asp Gln Glu Thr Leu Thr Glu Met Val Lys Pro Ser Ile
        280                 285                 290

gat tat gtg cgc cac aaa aaa ttc cga tct ggg aat tac cca tca tca     1088
Asp Tyr Val Arg His Lys Lys Phe Arg Ser Gly Asn Tyr Pro Ser Ser
    295                 300                 305

tta agc aat gaa aca gac cgg ctg gtg cac tgg tgc cac ggc gcc ccg     1136
Leu Ser Asn Glu Thr Asp Arg Leu Val His Trp Cys His Gly Ala Pro
310                 315                 320                 325

ggg gtc atc cac atg ctc atg cag gcg tac aag gtc ttt aag gag gag     1184
Gly Val Ile His Met Leu Met Gln Ala Tyr Lys Val Phe Lys Glu Glu
                330                 335                 340

aag tac ttg aaa gag gcc atg gag tgt agc gat gtg att tgg cag cga     1232
Lys Tyr Leu Lys Glu Ala Met Glu Cys Ser Asp Val Ile Trp Gln Arg
            345                 350                 355

ggt ttg ctg cgg aag ggc tac ggg ata tgc cat ggg act gct ggc cac     1280
Gly Leu Leu Arg Lys Gly Tyr Gly Ile Cys His Gly Thr Ala Gly His
        360                 365                 370

ggc tat tcc ttc ctg tcc ctt tac cgt ctc acg cag gat aag aag tac     1328
Gly Tyr Ser Phe Leu Ser Leu Tyr Arg Leu Thr Gln Asp Lys Lys Tyr
    375                 380                 385

ctc tac cga gct tgc aag ttt gca gag tgg tgt cta gat tac gga gca     1376
Leu Tyr Arg Ala Cys Lys Phe Ala Glu Trp Cys Leu Asp Tyr Gly Ala
390                 395                 400                 405

cac ggg tgc cgc att cct gac aga ccc tat tcg ctc ttt gaa ggc atg     1424
His Gly Cys Arg Ile Pro Asp Arg Pro Tyr Ser Leu Phe Glu Gly Met
                410                 415                 420

gct ggc gct att cac ttt ctc tct gat gtc ctg gga cca gag aca tca     1472
Ala Gly Ala Ile His Phe Leu Ser Asp Val Leu Gly Pro Glu Thr Ser
            425                 430                 435

cgg ttt cca gca ttt gaa ctt gac tct tcg aag agg gat taa              1514
Arg Phe Pro Ala Phe Glu Leu Asp Ser Ser Lys Arg Asp  *
        440                 445                 450

aaggtgcaaa aagacaacta aaatacccat ttggaccaaa agccgccaga ttgcttagtg   1574
cctgacacag aaacaactgg gaatcctgaa agagaagcag acaccgtcac aggcccctct   1634
ggttagacta gcatgagtga ccgaagccat ccatcaacat tttctaacag caccctcatc   1694
aatataaaat atgacttctt cacatacaaa aaaaaaaaaa aaagggcgg               1743
```

<210> 7
<211> 6

BNSDOCID: <WO___0149847A2_I_>

<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 7
Ser Ile Leu Thr Leu Thr
1 5

<210> 8
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 8
Ser Ile Leu Phe Leu Thr Cys
1 5

<210> 9
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 9
Asn Leu Tyr Ser Ser Ile Leu Phe Leu Thr Cys
1 5 10

<210> 10
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 10
Leu Ala Val Ala Asp Leu Leu
1 5

<210> 11
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 11
Leu Ala Leu Leu Leu Thr
1 5

<210> 12
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 12

Leu Arg Arg Ser Leu Pro
1 5

<210> 13
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 13
Phe Leu Val Gly Asp Pro Gly Asn Ala
1 5

<210> 14
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 14
Gly Asn Ala Met Val
1 5

<210> 15
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 15
Leu Ala Val Ala Asp
1 5

<210> 16
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 16
Phe Leu Val Gly Val Pro Gly Asn Ala
1 5

<210> 17
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 17
Ala Leu Leu Leu Thr
1 5

<210> 18
<211> 10
<212> PRT
<213> Artificial Sequence

BNSDOCID: <WO 0149847A2_I_>

<220>
<223> amino acid fragment

<400> 18
Ala Asp Leu Leu Cys Cys Leu Ser Leu Pro
1 5 10

<210> 19
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 19
Tyr Val Gly Ala Ala His Gly
1 5

<210> 20
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 20
Leu Val His Trp Cys His Gly Ala Pro Gly Val Ile
1 5 10

<210> 21
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 21
Gln Ala Tyr Lys Val Phe
1 5

<210> 22
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 22
Glu Glu Lys Tyr Leu
1 5

<210> 23
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 23
Ser Leu Phe Glu Gly Met Ala Gly
1 5

<210> 24
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 24
Arg Phe Pro Ala Phe Glu Leu
1 5

<210> 25
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 25
Leu Leu Gln Gln Met Glu
1 5

<210> 26
<211> 12
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 26
Thr Phe Leu Cys Gly Asp Ala Gly Pro Leu Ala Val
1 5 10

<210> 27
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 27
Ala Gly Ile Tyr Tyr
1 5

<210> 28
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 28
Ser Gly Asn Tyr Pro
1 5

<210> 29
<211> 9
<212> PRT
<213> Artificial Sequence

<220>

PNSDOCID: WO 0149847A2 | >

<223> amino acid fragment

<400> 29
Gln Ala Tyr Lys Val Phe Lys Glu Glu
1 5

<210> 30
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 30
Asp Val Ile Trp Gln
1 5

<210> 31
<211> 17
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 31
Lys Tyr Leu Tyr Arg Ala Cys Lys Phe Ala Glu Trp Cys Leu Asp Tyr
1 5 10 15
Gly

<210> 32
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 32
Glu Leu Leu Tyr Gly Arg
1 5

<210> 33
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 33
Pro Tyr Ser Leu Phe Glu Gly
1 5

<210> 34
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> amino acid fragment

<400> 34
Val Thr Phe Leu Cys Gly
1 5